The Role of Pollen Tube Reception in Reproductive Isolation

Dissertation zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Zürich von

Lena Maria Müller

aus Deutschland

Promotionskomitee

Prof. Dr. Ueli Grossniklaus (Vorsitz und Leitung der Dissertation)

Prof. Dr. Beat Keller

Prof. Dr. Karl Schmid

Zürich, 2014

Acknowledgements

First of all, I want to thank **Ueli Grossniklaus** for giving me the opportunity to do my PhD in his lab and work on an exciting topic, furthermore I want to thank him for his support, guidance, and for his great ideas.

I want to thank **Beat Keller** (University of Zurich) and **Karl Schmid** (University of Hohenheim) for being part of my thesis committee, for reading and evaluating my work and for being co-examiners at my PhD defense.

A very big thank you goes to **Heike Linder** for hours of great discussions, fun times while doing experiments and of course for being a such great friend also outside the lab.

I am very grateful to **Nuno Pires** for all the help with statistics and mapping, planting for hours in the greenhouse and always having a good time!

I also want to thank **Marc Schmid** for discussions and help concerning next-generation sequencing and also for always being up for a beer.

Furthermore, I want to thank all the present and former members of the Sex Club (**Heike, Anna, Moritz, Nadine, Christina, Evelyne, Aurel and Sharon**) for very helpful discussions every other week and for all the fun we had.

I am very grateful to all technical staff (**Arturo** for ordering, **Christof** for microscopy and **Peter** for saving us a lot of time by supplying us with solutions). A special thanks is for **Valeria** for her help with digital PCR and lots of other things during daily lab life. And, last but not least, of course also to **Daniela** for sequencing, help with emasculations and always being there for me!

I want to thank all the girls in my lab (P2-27) for creating a great working atmosphere: **Milka**, **Anna**, **Klara**, **Daniela**, **Nadine**, **Diana** and of course also the "lost boys" **Christian H**., **Mayank** and **Aurel**.

Of course, working in the lab would not be as nice without all the great people creating such a nice working atmosphere and became really good friends: **Heike, Debo, Daniela, Evelyne, Michi, Aurel, Sharon, Kostas, Moritz, Klara, Anja H., Nadine, Milka, Anna, Christian S. and H., Stefan G., Marc, Nuno, Sofia, Joana, Nina, Diana, Polly, Christina, Valeria, Arturo, Christof, Fred, Sam, Matthias, Celia, Kinga, Jasmin, Wenjing, Bruno, Roger, Afif, Hannes, Wanhui, Guillaume, Michael, Anja S., Önder, Dima and Dima, Ulrike, Quy, Sibylle, Johan, Marek, Marian, Bobi, Mayank, Stefan W., Hampe and Annegret.**

Last but not least, I want to thank **my family and friends** for all their moral support during the last years and for taking my mind off things. I especially want to thank **Oli** for all his love and support, and for always being there for me.

Zusammenfassung

Arten sind voneinander durch verschiedene Artgrenzen isoliert, welche den Genaustausch verhindern. Bei Pflanzen können zum Beispiel geographische oder zeitliche Faktoren wie unterschiedliche Blühzeitpunkte, Habitate oder Bestäuber verhindern, dass Pollen von einer Art auf eine andere übertragen wird. Wenn doch einmal eine Fremdbestäubung stattfindet, kann das Fruchtblatt durch Kommunikation mit dem Pollen feststellen, ob dieser zur gleichen Art gehört und artfremden Pollen gegebenenfalls abweisen. Dies gewährleistet, dass ausschließlich die Keimzellen kompatibler Pollen die weiblichen Gameten (Ei- und Zentralzelle) befruchten. Dadurch wird die Entstehung hybrider Nachkommen verhindert, welche oft nicht lebensfähig oder steril sind. In Kreuzungen zwischen den nah verwandten Brassicaceae-Arten *Arabidopsis thaliana* und *A. lyrata* erfolgt die Erkennung des Pollens zum Zeitpunkt des Pollenschlauchempfangs. So bezeichnet man den ersten direkten Kontakt des männlichen Gametophyten (der die Keimzellen transportierende Pollenschlauch) mit dem weiblichen (der tief in das Fruchtblatt eingebettete Embryosack). Wenn ein *A. lyrata* Pollenschlauch nicht von einem weiblichen *A. thaliana* Gametophyten erkannt wird, so empfängt dieser kein Signal um seine Keimzellen freizusetzen, sondern wächst weiter in den Embryosack hinein (Pollenschlauchwucherung). Da die betroffenen weiblichen Gametophyten nicht befruchtet werden, glauben wir, dass eine solche Pollenschlauchwucherung eine Artgrenze darstellt.

Das Ziel dieser Dissertation war es, die genetische Basis des Pollenschlauch-Erkennens zu untersuchen. Da Mutanten in der Rezeptor-ähnlichen Kinase *feronia* ebenfalls Pollenschlauchwucherungen zeigen, untersuchten wir als erstes, ob das Kandidaten-Gen *FERONIA* direkt an der Erkennung artfremder Pollenschläuche beteiligt ist. Dies scheint jedoch nicht der Fall zu sein. Also wollten wir neue Gene identifizieren, die an diesem Prozess mitwirken und nutzten dazu die bemerkenswerte natürliche Variation in *A. thaliana* Akzessionen. Wir haben 86 Akzessionen untersucht und solche gefunden, welche nur in 10% ihrer weiblichen Gametophyten *A. lyrata*-Pollenschlauchwucherung aufweisen, und solche, in denen mehr als 90% betroffen sind. Mit Hilfe einer Genom-weiten Assoziations-Kartierung konnten wir zeigen, dass eine Untereinheit der pflanzlichen Oligosaccharyltransferase an der Unterscheidung von artfremden und arteigenen Pollenschläuchen beteiligt ist. Dies könnte bedeuten, dass artspezifische Glykosylierungsmuster die Pollenschlaucherkennung beeinflussen. Zudem nutzen wir die natürliche Variation weniger Akzessionen mit entweder sehr niedrigen oder sehr hohen Anteilen an Pollenschlauchwucherung. Aus Akzessionen mit extremen Phänotypen wurden segregierende F_2 Kartierungspopulationen hergestellt. Segreganten wurden mit Hilfe modernster Sequenzierungsmethoden sequenziert und so konnten wir ein kleines Peptid mit einer Glykan-Bindedomäne identifizieren, das an der Erkennung artfremder Pollenschläuche beteiligt ist. Dies verdeutlicht noch einmal die immense Bedeutung von Saccharid-Signalen für den artspezifischen Pollenschlauchempfang.

Abstract

Species are reproductively isolated by a variety of speciation barriers restricting gene flow between them. In plants, these barriers can be spatial or temporal patterns preventing pollen transfer from one species to another, for example differences in flowering time, habitat or pollinator. However, if interspecific pollen transfer has occurred, extensive communication between the pollen and the female floral tissues allow the female to recognize and reject interspecific pollen. This ensures that only sperm cells from compatible pollen succeed to fertilize the plant's female gametes, the egg and the central cell (double fertilization), thereby preventing the formation of potentially unviable or sterile hybrid offspring. Between the closely related Brassicaceae species *Arabidopsis thaliana* and *A. lyrata*, pollen recognition by the female occurs during pollen tube reception, a process constituting the first direct contact between the male gametophyte (the tip-growing pollen tube transporting the sperm cells) and the female gametophyte (the embryo sac, which is deeply embedded in sporophytic tissue). If an *A. lyrata* pollen tube is not recognized by an *A. thaliana* female gametophyte, it does not receive a female signal to release its sperm cells and instead invades the embryo sac (pollen tube overgrowth), leaving the embryo sac unfertilized. Therefore we believe that pollen tube overgrowth constitutes a species barrier because no interspecific hybridization event occurs.

The aim of this thesis was to examine the genetic basis underlying the pollen tube recognition process. Because mutants in the receptor-like kinase *feronia* exhibit a similar pollen tube overgrowth phenotype as observed in interspecific crosses, we first conducted a candidate gene approach to elucidate the function of *FERONIA* in interspecific pollen tube recognition. However, FERONIA does not seem to be directly regulating this process. Second, we aimed to identify novel factors involved in pollen tube recognition exploiting the striking natural variation of *A. thaliana* accessions in interspecific pollen tube reception: we found that 86 *A. thaliana* accessions vary in their ability to recognize *A. lyrata* pollen tubes, displaying between 10% and 90% of ovules with pollen tube overgrowth in a silique. We used genome-wide association mapping to correlate the phenotypic with the genotypic variation and identified a subunit of the plant oligosaccharyltransferase complex to be regulating interspecific pollen tube reception. This indicates that species-specific glycosylation patterns might be involved in establishing species barriers during pollen tube reception. Furthermore, we made use of the natural variation in pairs of *A. thaliana* accessions showing very high and very low proportions of interspecific pollen tube overgrowth. By crossing accessions with extreme phenotypes, we created segregating F_2 mapping populations. With the help of Next-Generation Sequencing of bulks of these segregants, we were able to identify a small peptide with a glycan-binding domain to be a regulator of the recognition of inter- and intraspecific pollen tubes. This finding further strengthens the significance of carbohydrate patterns during pollen tube reception.

Contents

Scope of the thesis

Pollen tube reception is an important species-recognition step during plant fertilization, because it enables the female gametophyte (embryo sac) to directly communicate with the male gametophyte (pollen tube). This cross-talk allows the embryo sac to discriminate between a pollen tube from the same or from another species, allowing the rejection of interspecific pollen tubes and thus preventing hybrid formation.

The aim of this thesis was to identify molecular components regulating the recognition of pollen tubes by the female gametophyte. In order to identify such factors, we conducted a candidate gene approach with the receptor-like kinase *FERONIA* (chapter 2). In addition, we were able to identify novel genes involved in interspecific pollen tube reception making use of natural variation in *A. thaliana* (chapters 3 and 4).

Every chapter of this thesis is written in the form of a publication, containing a separate introduction specific to the content, results, methods and discussion as well as all references cited in the respective chapter. Chapter 1 comprises a general introduction to the field of plant reproduction, speciation as well as cell-cell communication and a published review about the *Catharanthus roseus* receptor-like serine/threonine kinase1-like (*CrRLK1L*) family. In chapter 5, a general discussion about the results of this thesis is presented and the findings are put in a general context of reproductive isolation and plant fertilization. The appendix includes vector maps and larger tables.

Chapter 1 – Introduction

1 – Speciation in Plants

Classically, a species is defined as a population of individuals that can freely interbreed among each other, but are reproductively isolated from individuals of another species. The formation of plant species, referred to as speciation, is a process that drives the evolution of within-species variation in populations to the establishment of taxonomically different species by building up hybridization barriers (*1*). These hybridization barriers are preventing gene flow between species and can be divided into pre- and post-pollination barriers, whereas the latter can be further subdivided into pre- and postzygotic hybridization boundaries (Fig. 1.1, (*2*)), acting before and after zygote formation, respectively. Pre-pollination barriers have strong individual isolation strengths, hence they alone can provide strong reproductive barriers (*3*). However, many species barriers are established by a combination of different pre- and post-pollination mechanisms.

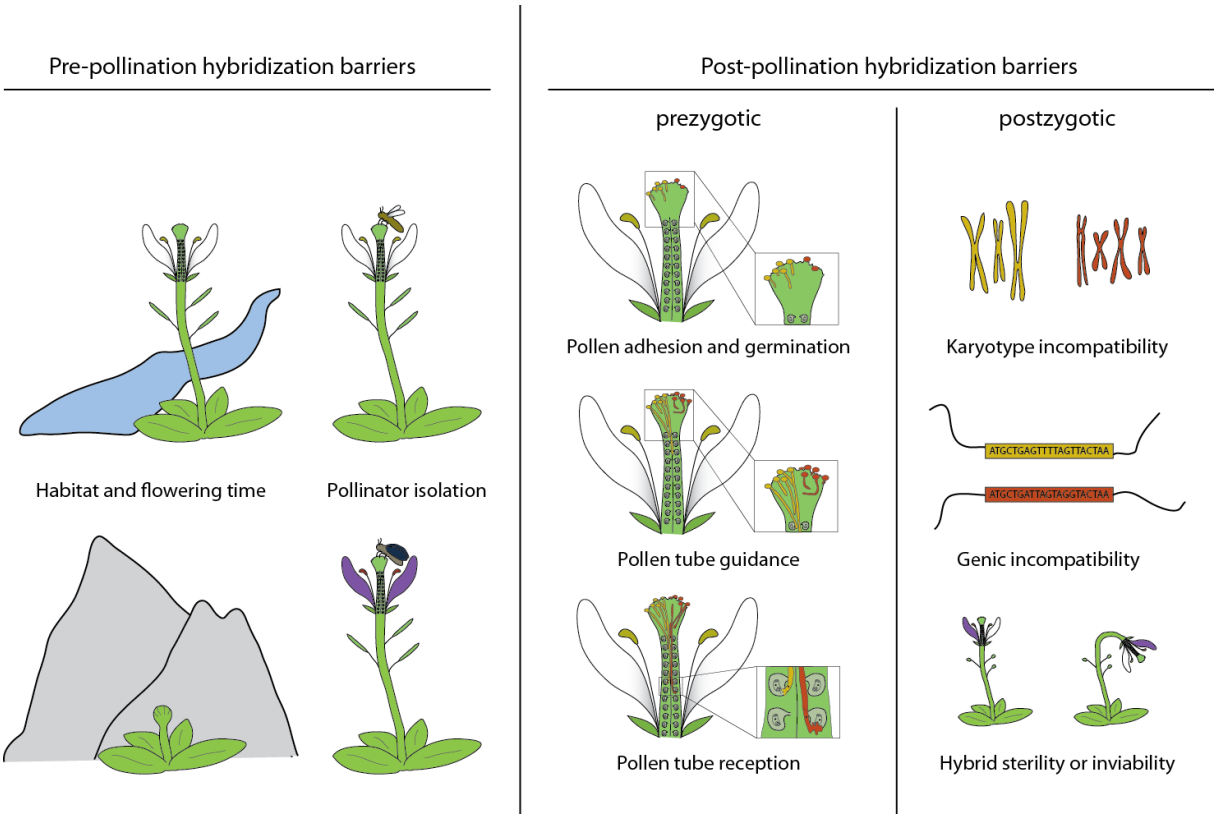


Figure 1.1: Species are separated by multiple layers of hybridization barriers. Pre-pollination barriers prevent interspecific pollen transfer. Post-pollination, prezygotic barriers act during pollen adhesion, pollen tube germination and growth, as well as pollen tube reception. Post-pollination, postzygotic barriers are genic or genomic incompatibilities leading to hybrid sterility or inviability. Figure adapted from (*2*).

Genes determining hybridization barriers and therefore contributing to reproductive isolation are called speciation genes. These genes often display signatures of positive Darwinian selection, however most evidence is from the animal field and only little is known in plants (*4*, *5*). Positive selection is characterized by an excess of non-synonymous substitutions on the protein level due to fixation of beneficial mutations and thus is evidence for rapid adaptive evolution, which can be caused for instance by a molecular arms race between genes involved in hybrid incompatibilities (*6*). A molecular arms race can be triggered by intrinsic selfish elements driving the evolution of their "restoring" counterparts, for example during cytoplasmic male sterility, (see section 1.3, (*7*)) or by selective pressure in plant-pathogen interactions. Indeed, some genes involved in pathogen response cause hybrid incompatibilities (see section 1.3, (*8*)). In animals, sexual selection, such as mate choice, sperm competition and sperm-egg interaction, leads to rapid evolution of genes involved in those processes. For example, proteins conferring species-specific interactions of gametes, such as the abalone sperm protein lysin and its egg coat receptor VERL, have been found to co-evolve rapidly under positive selection (*9*). In *Arabidopsis*, male gametophyte-specific genes (expressed in the sperm cells, pollen and pollen tube) show a higher proportion of adaptive amino acid substitutions compared to female gametophytic-specific genes and a set of randomly chosen sporophytic genes, presumably caused by increased purifying and positive selection in the pollen-specific genes (*10*). These findings indicate that sexual selection might also occur in plants, possibly through interactions during pollen tube growth leading to pollen tube competition.

In the following sections, I will describe the different mechanisms of reproductive isolation barriers in plants with a focus on the underlying genetic basis.

1.1 – Pre-Pollination Hybridization Barriers

Pre-pollination barriers are temporal or spatial factors that prevent species from cross-pollinating. This includes (i) habitat isolation, (ii) flowering time, (iii) pollinator isolation, and (iv) mating system divergence.

(i) Habitat isolation is the first step of allopatric speciation diminishing gene flow between geographically separated populations (ecotypes) (*6*, *11*). Adaptive mutations to the new environment will strengthen habitat isolation, thus leading to almost complete

reproductive isolation. If these allopatric populations further evolve genetic characteristics that prevent genetic interchange between them, the formation of a new species is complete (*2*). Although habitat isolation is among of the strongest speciation barriers, evidence for genes involved in the formation of this barrier is still missing (*3*). One example for adaptive evolution in the genus *Arabidopsis* is the heavy-metal tolerance of *A. halleri*, as opposed to the related species *A. lyrata* and *A. thaliana*. In *A. halleri*, the metal transporters *MTP1* and *HMA4* are strongly up-regulated compared to its metal-sensitive relatives, allowing the species to tolerate soil with high heavy metal content (*12*, *13*). Although these adaptations are not establishing actual hybridization boundaries because they do not interfere with reproduction *per se*, they still contribute to geographic isolation, which allows other genetic isolation barriers within the *Arabidopsis* genus to evolve (*2*).

(ii) Especially for outcrossing species, synchronized flowering is critical for reproductive success as it ensures the presence of potential mating partners. Many studies have addressed intraspecific natural variation of flowering time in a number of species (*14-16*), however, the molecular basis of flowering time as an interspecific hybridization barrier is poorly understood. A recent study identified QTLs for flowering time differences, a major reproductive isolation barrier between two *Mimulus* species that are able to hybridize in the greenhouse (*17*).

(iii) Major changes in floral traits are often accompanied with altered pollinator attraction, which in turn can be leading to species isolation. Among the floral traits that strongly influence pollinator preference are olfactory cues, visual cues such as flower color, and mechanical attributes like texture, floral tube length or shape (*18*). Although pollinator-driven reproductive isolation is one of the strongest factors in speciation, its genetic basis is largely unknown (*3*). Among the genes that were identified to be crucial factors in plant-pollinator interaction is a set of stearoyl-acyl carrier protein desaturases. The expression level differences of these genes determine alkene composition in the odor of reproductively isolated sexually deceptive *Ophrys* species, thus determining specificity of pollinator attraction (*19*). Most known genes involved in flower color determination belong to the group of *MYB*-transcription factors that are regulating the anthocyanin biosynthesis or key enzymes in this pathway (*20-22*). For example in *Petunia*, it was shown that the *MYB*-transcription factor *ANTOCYANIN2* is a

key determinant of flower color variation between *P. integrifolia* and *P. axiliaris*, which is going along with the attraction of different pollinator species (*20*).

(iv) Mating-system differences constitute the strongest group of reproductive isolation barriers (*3*) and thus an extreme case of pollinator-mediated reproductive isolation is the evolution of selfing (autogamy), which can be the result of an absence of suitable pollinators (*23*). The evolution of selfing requires the loss of self-incompatibility and selfing – although often leading to inbreeding depression – is considered to be advantageous when colonizing new territory (*24*). Besides the loss of self-incompatibility, another important step is the promotion of selfing rather than outcrossing. The autogamous *Solanum lycopersicum* has recessed stigmas compared to its outcrossing relatives with exerted stigmas. This change in floral morphology causes predominant self-pollination, and is caused by down-regulation of a single transcription factor, *STYLE2.1* (*25*).

Taken together, pre-pollination barriers display a wide range of mechanisms but only few underlying genes have been identified. Pre-pollination barriers are the strongest forces maintaining species identities, because they are the first to occur and each subsequent barrier can only prevent the gene flow that has not yet been blocked by prior barriers. However, during the evolution of reproductive isolation, it is possible that post-pollination barriers evolved first and were the actual cause for the formation of a new species (*26*).

1.2 – Post-Pollination, Prezygotic Hybridization Barriers

Post-pollination, prezygotic hybridization barriers act after pollination, but before the formation of the zygote (*27*). These barriers strongly rely on direct interactions of the male gametophyte (pollen) with the female tissues in order to determine species-specificity and ensure fertilization success: during pollen hydration on the stigma, pollen tube growth through the transmitting tract and guidance towards the ovules (*28*). In the end, the male and the female gametophyte (embryo sac) as well as the gametes directly communicate with each other during pollen tube reception and gamete fusion, respectively. Post-pollination, prezygotic barriers can be further subdivided into two groups: (i) conspecific pollen precedence (CPP), where conspecific pollen tubes have an advantage compared to heterospecific pollen tubes (male competition); and (ii) gametic

or gametophytic incompatibilities, in which heterospecific pollen tube fail to fertilize egg cells (female preference) (*27*).

(i) Competition between pollen tubes in the style is a major reproductive isolation barrier if pollen from different species with overlapping pollinators and geographical distribution are frequently deposited on a plant's stigma. The phenomenon of pollen competition has been studied in a number of species, including *Mimulus*, *Iris*, *Lilium*, *Oryza* and *Silene* and usually results in fewer hybrid seeds than expected after pollinating the stigmas with a mixture of con- and heterospecific pollen (*29*-*33*). In interspecific crosses within the genus *Mimulus*, pollen tubes of *M. guttatus* outcompete those of *M. nasutus* in the *M. guttatus* style due to higher pollen tube growth rates (*29*). In a recent shotgun proteomic approach of pollen tubes within *M. guttatus* styles where CPP occurs, a number of candidate genes that were pollen-expressed and under positive natural selection was found, although none of them has been experimentally verified yet (*34*).

(ii) In contrast to the limited knowledge about the molecular mechanisms underlying pollen competition, several genes involved in interspecific gametophytic incompatibilities have been isolated, for example the *Nicotiana S*-RNase as well as other self-incompatibility factors, and pollen tube attractants LUREs in *A. thaliana* and *Torenia* ((*35*-*37*); see section 2). These barriers usually rely on incongruity based on the divergent evolution of reproductive signals rather than a general inhibition of interspecific pollen (*38*).

Because this work mainly focuses on the identification of gametophytic hybridization barriers, their molecular basis will be discussed in more detail in section 2.

1.3 – Post-Pollination, Postzygotic Hybridization Barriers

Postzygotic barriers are selecting against hybrid offspring of an interspecies cross on the genic (i) or genomic (ii) level, causing inviability or sterility of the hybrid offspring. This selection can occur either in the first or in subsequent hybrid generations and either at the embryonic or adult stage.

(i) Postzygotic barriers on the genic level are caused by epistatic interactions of two alleles originating from the two parental species at two or more loci (*39*-*42*). The Bateson-Dobzhansky-Muller model indicates that such "complementary genes" and their

alleles usually perform well in an intraspecies environment but cause problems in hybrids, when combined with interspecific alleles they would normally never encounter. The initial genetic experiments that lead to the elaboration of the Bateson-Dobzhansky-Muller model were conducted in *Drosophila* (*40*, *41*), but also in plants, epistatic interactions of genes leading to hybrid necrosis (also called hybrid weakness) have been identified (*8*). Hybrid necrosis resembles in many aspects the immune response to pathogen attack, with phenotypes such as cell death, tissue necrosis, yellowing, chlorosis and dwarfism. These conditions, which can occur in intra- and interspecies hybrid F_1 or F_2 generations, are usually lethal or lead to sterility and were observed in in a variety of plant species both in laboratory crosses and in the wild (reviewed in (*43*)). Hybrid necrosis and pathogen response share a common basis: both involve increased oxidative stress, elevated levels of pathogen response genes, and programmed cell death; and several cases of hybrid necrosis in intraspecific combinations of *A. thaliana* accessions involve known disease resistance genes (*8*, *44*). Based on these findings it has been hypothesized that the pathogen-derived selection pressure has shaped the basis for epistatic interactions that cause intra- and interspecific hybrid incompatibility (*8*).

A special case of Bateson-Dobzhansky-Muller incompatibilities is cytoplasmic male sterility (CMS), a plant-specific phenomenon that affects male gametophyte development and is characterized by the complete absence of pollen or anthers. CMS often arises in hybrids (alloplasmic CMS) and is caused by incompatibilities between mitochondrial and nuclear genes originating from different parental origins (*45*). Because mitochondria are maternally inherited and the nuclear genome is a mix of maternal and paternal contributions, a genic conflict between the genomes can arise. Usually, CMS is caused by rearrangements of the mitochondrial genome, giving rise to chimeric open reading frames coding for selfish elements that consist of mitochondrial genes and unknown reading frames. Most of these chimeric genes involve ATP synthase genes or subunits of the cytochrome oxidase and NADH oxidase that often have cytotoxic effects (reviewed in (*45*)). These cytotoxic effects can be restored by nuclear restorer of fertility (*RF*) genes, most of which belong to the pentatricopeptide repeat (PPR) family (*46*). Thus, interspecific hybrids inheriting only the deleterious mitochondrial CMS allele but not the nuclear RF gene might be male-sterile and therefore, CMS is likely a cause for interspecific hybridization barriers (*43*).

(ii) Next to genic incompatibilities, karyotype incompatibilities between the parental species can cause a failure in meiosis in the heterozygote hybrid generation by preventing proper chromosome pairing, thus leading to sterility of otherwise viable plants (*27*). Chromosomal rearrangements are another possible cause for hybrid incompatibilities in plants, a phenomenon that arises by chromosomal recombination during meiosis of the heterozygote hybrid. Translocations between (interspecific) homologous chromosomes can lead to unbalanced chromosomal content in the haploid generation, leading to unviable gametes (*27*, *47*). However, it is unknown whether chromosomal rearrangements have an intrinsic, direct effect on hybrid viability and sterility, or if they promote accumulation of genes causing Bateson-Dobzhansky-Muller incompatibilities.

2 – Species-Specific Signals during the Fertilization Process

Fertilization describes the process which begins with the deposition of the pollen grains onto the stigma of the carpel and ends with double fertilization of the haploid egg- and the diploid central cell by the two male sperm cells, giving rise to the embryo and its nourishing tissue, the triploid endosperm.

In many species, the pollen is a tri-cellular structure that develops in the anthers after meiotic division of the microspore mother cell. This asymmetric division gives rise of a large vegetative cell and a smaller generative cell, which further divides mitotically to produce two sperm cells (*48*). The three cells of the pollen are packed in several layers of cell wall: the inner cell wall (intine), which resembles the general composition of primary plant cell wall consisting of cellulose, hemicellulose, pectin and proteins; and the outer cell wall (exine), which is composed of hydrophobic sporopollenin. Sporopollenin is sculpted in taxon-specific microscopic patterns and is composed of a so far not exactly determined mixture of biopolymers, making it highly resistant to biological, physical and chemical agents (*49*). The outermost layer of the pollen grain consists of the pollen coat, which is to a great extent composed of lipids but also contains considerable amounts of proteins, most of which belong to the family of oleopollenins (or oleosins) (*50*). The extracellular pollen coat protects the pollen from dehydration, but is also crucial for mediating pollen-pistil interactions (see section 2.1). In Angiosperms, the female gametophytes (embryo sacs) are deeply embedded in pistil

tissue, making it necessary for the pollen to transport the immobile sperm cells with the help of a long, tip-growing vegetative cell (the pollen tube, PT) towards the ovules that harbor the embryo sacs. In the PT, the two interconnected sperm cells are transported in a so called “male sperm unit” together with the vegetative nucleus (*51*).

The multicellular embryo sac of angiosperms is embedded in sporophytic ovule tissue and is composed of four cell-types: the two synergid cells, which are located at the micropylar opening of the ovule where the pollen tube enters; two gametes (one egg and one central cell); and three antipodal cells at the chalazal end of the ovule. All cells of the female gametophyte are derived from a single megaspore mother cell, which undergoes meiosis, followed by syncytial nuclear divisions giving rise to an eight-nucleate embryo sac (*52*). During cellularization of the gametophyte, the eight nuclei are distributed into the four above mentioned cell types. All of these cells are haploid, except for the diploid central cell, which is the result of the fusion of two polar nuclei.

Plant fertilization can be subdivided into several stages: i) adhesion and hydration of the pollen on the stigma, ii) germination and growth of the PT, iii) PT reception, and iv) the fusion of the gametes during double fertilization. All these processes require communication between the male gametophyte and the cells of the female tissues such as the stigma (in stage i and ii), the style and transmitting tract (in stage ii) and the cells of the embryo sac (in stages ii,iii and iv). Many of the molecular cues that mediate the communication between male and female tissues are known and there is evidence that at least some of the factors mediating pollen-pistil interactions are able to discriminate between pollen of other species, ensuring that only intraspecific PTs are able to arrive at the embryo sacs and deliver their sperm cells (see below). Thus, the fertilization process is largely under female control (*53*).

The following sections describe the processes of pollen hydration, PT germination, guidance and growth, PT reception and sperm-egg fusion in greater detail with a focus on the molecular basis of species-specific interactions between male and female tissues. However, genes that have been implicated in processes like general PT growth, gametogenesis as well as gametophyte and pistil development and are therefore only indirectly influencing male-female interactions, are beyond the scope of this thesis and will not be covered in this introduction.

2.1 – Pollen Grain Adhesion and Hydration on the Stigma

Pollen grains can be deposited on the stigma by wind, water or pollinating animals. Thus, the first step ensuring successful fertilization is the proper attachment of the pollen grain to the stigmatic papillae cells, a process known as pollen capture. Pollen capture is only dependent on the exine, whereas the pollen coat seems not to be involved during this initial step. This could be shown by the impaired pollen adhesion of various mutants affecting sporopollenin biogenesis and by the fact that purified exine alone is able to adhere to the stigma (*54*, *55*). In a detergent wash assay, it was further shown that *A. thaliana* pollen or even the exine only adheres stronger to *A. thaliana* stigmas than to stigmas of other Brassicaceae (*54*). This species-specific pollen-stigma adhesion roughly correlates with the distance of taxa, since monocot pollen show a generally weaker adherence to *A. thaliana* stigmas than pollen from dicotyledons. The molecular basis of specific pollen capture is unknown, but sporopollenin composition in the exine varies along taxa, thus harboring a great potential for species-discrimination (*28*).

Plant stigmas can be of two types, which determine the mechanism of pollen capture: wet stigmas (in species such as *Lilium, Nicotiana* and *Petunia*) secrete an exudate consisting of water, proteins, lipids and carbohydrates. These stigmas non-selectively capture pollen regardless of their origin by liquid surface tension, indicating that in species with wet stigmas, species-specific recognition of male and female reproductive tissues only occurs after pollen adhesion and hydration (*56*). In contrast, dry stigmas (such as the ones from *Zea mays* and *Arabidopsis*) do not deposit a liquid exudate on their surface and thus, adhesion and hydration are highly regulated and allow species-discrimination. Dry stigmas consist of protruded epidermal cells (papillae) that are covered with a cuticle highly permeable for water and larger molecules. On top of the cuticle sits a thin, waxy protein layer (pellicle) consisting of esterases and glycoproteins (*38*). After the initial capture of the pollen grain, the pollen coat gets mobilized, floats toward the papilla cells and creates an interface with the papilla cells (a “foot”). In this interface, lipids, carbohydrates and proteins from the male and female reproductive organs mix for the first time, and thus provide opportunities for molecular species-specific recognition events. In *Brassica*, two potential receptor-ligand pairs that are expressed in the stigma cells and the pollen, respectively, have been proposed to mediate the crosstalk leading to “foot” formation of compatible pollen grains (pollen

adhesion): *SLR1* (*S-LOCUS RELATED1*) encodes a stigmatic receptor and is not part of the *S*-locus, but displays high sequence similarity to *SLG* (*S-LOCUS GLYCOPROTEIN*) and *SRK* (*S-LOCUS RECEPTOR KINASE*) (*57*). SLR1 was shown to bind to the pollen coat proteins (PCPs) SLR-BINDING PROTEIN 1 and 2 (SLR-BP1 and SLR-BP2) (*58*), however, functional validation for the involvement of SLR-BP1 and 2 in pollen adhesion is still missing. Masking SLG with monoclonal antibodies on the stigma of *B. oleracea* resulted in reduced pollen adhesion ability (*59*). Like SLR1, SLG also binds to a member of the PCP family, PCP-A1, but genetic evidence for a direct involvement of this receptor-ligand pair in pollen adhesion is lacking (*60*). Although successful pollen adhesion does not seem to be species-discriminant within the Brassicaceae, adhesion of more distantly related species on *Brassica* stigmas is impaired (*61*). Thus, it is possible that sequence variants of the receptor-ligand pairs SLR1/SLR-BP1,2 and SLG/PCP-A1 could be candidates for determining species-specific pollen adherence.

Pollen is released from the anthers as dehydrated microspores and hydration is crucial in order to activate the metabolism necessary for PT germination. In *Arabidopsis*, compatible pollen is hydrated 5 minutes after successful adhesion on the stigmatic cells, whereas incompatible pollen derived from species outside the crucifer family does not hydrate (*54*, *62*). During pollen hydration, the stigma serves as the source of water, and lipids from both the female cuticle and the pollen coat were found to be essential (*38*). Mentor pollen experiments with compatible and incompatible pollen grains revealed that simple indirect water transfer from one pollen grain to another can be excluded during hydration, as it was shown that direct contact of the pollen grain with the stigma surface is crucial for hydration (*62*). The molecular basis underlying pollen hydration has long remained elusive until recently the stigmatic protein EXO70A1 has been shown to influence pollen hydration in *Brassica* and *Arabidopsis* (*63*). EXO70A1 is a putative member of the exocyst complex regulating polarized secretion and was identified as a member of the self-incompatibility pathway in *B. napus*. Loss of *EXO70A1* function in the self-compatible *Arabidopsis* mimicked *Brassica* SI response and overexpression in self-incompatible *Brassica* lead to self-pollen acceptance (*63*). The exocyst complex is proposed to dock vesicles to phosphatidolinositol-phosphate-residues (PIP) in the plasma membrane, thus promoting vesicle fusion and secretion. Consequently, mutants in the PIP biosynthetic pathway show delayed pollen hydration on the stigma (*64*). On the pollen side, a family of glycine-rich oleosin proteins (GRPs) was implied in pollen

hydration. Mutations in the most abundant pollen coat GRP, *grp17*, showed reduced pollen hydration, and phylogenetic analysis of the gene family in closely related species of the Brassicaceae revealed rapid evolution of the genes through positive selection (*65*, *66*).

2.2 – Pollen Tube Germination and Growth in the Pistil

In *Arabidopsis*, 15 minutes after the pollen grains have adhered to the stigma, PT germination commences (*54*). The tip-growing PT proceeds through the "foot" established by the pollen coat (see section 2.1), elongates and subsequently penetrates the cell walls of the stigmatic papillae cells. When it reaches the base of the papillae, the PT continues growing intercellularly through the style towards the transmitting tissue of the septum. Once the PT leaves the transmitting tract, it grows along the funiculus towards the micropylar end of the ovule, where the journey of the PT ends with its rupture and the release of the sperm cells (discussed in section 2.3, (*67*)).

Mainly *in vitro*, many studies have addressed the internal signaling pathways and cellular processes involved in PT germination and growth (including cell wall remodeling, calcium signaling, reactive oxygen species, phospholipids, cyclic AMP, protein kinases, NADPH oxidases and GTPases) (reviewed in (*68*)); however, less is known on the interactions of the PT with the pistil. *In vivo*, PTs display a higher growth rate than *in vitro* and in line with that, pistil-growing PTs have a different transcriptome than *in vitro*-growing ones, suggesting that pollen-pistil interactions strongly influence PT growth (*69*, *70*). In the following section, I will not describe the basic signaling mechanisms during (*in vitro*) PT elongation, but only concentrate on the findings that concern the pollen-pistil interactions during PT germination and growth.

2.2.1 – Pollen Tube Germination

On the stigma, the hydrated pollen grain establishes a polarity by re-organization of its cytoplasm and cytoskeleton, a process mediated by the actin-bundling protein *FIMBRIN5*, which cross-links and stabilizes actin filaments (*71*, *72*). In addition to changes in the cytoskeleton, callose is transported towards the site of PT emergence and introduced into the cell wall, giving rise to the collar-like annulus, a pore at the base of the emerging PT. An overall increase in the levels of cell wall pectins and

arabinogalactan proteins (AGPs) was observed, which are secreted to the cell wall at the pollen aperture and might play roles in stabilizing the emerging PT as well as in recognition processes between the PT and the papillae cells (*73*).

Similar to pollen hydration, PT germination also strongly relies on stigmatic signals, and thus the exocyst complex subunit EXO70A1 is also crucial for PT germination and penetration of the stigma (*63*). The exact nature of the pistil factors promoting PT germination on the stigma is not known, but it is thought that the pistil provides enzymes to modify the cuticle and cell wall of papillae cells, making them accessible for the invading PT (*67*, *74*). Calcium (Ca^{2+}) signaling is fundamental for PT growth and germination (*75*). Internal oscillating Ca^{2+} gradients in growing PTs regulate vesicle trafficking and cell wall remodeling required for tip-growth, but also external Ca^{2+} from female tissues is essential. In species with wet stigmas, the secreted liquid on the stigma surface contains Ca^{2+} and it is believed that already during pollen hydration, Ca^{2+} enters the pollen grain through the developing germinal pore (*76*). Later, during PT growth through the stigma and the transmitting tract, Ca^{2+} precipitates potentially originating from stigma/pistil vesicles are visible at the pollen tube tip. As it could be shown with the Ca^{2+} sensor yellow cameleon in *Arabidopsis*, papillae cells and pollen seem to communicate via Ca^{2+} oscillations: On the female side, three peaks of cytosolic free Ca^{2+} were observed just at the site of pollen grain adhesion in the stigmatic papillae cells: after pollen hydration, protrusion and when the PT penetrated the papillae (*77*). These female Ca^{2+} maxima correlated with the behavior of the pollen, since shortly after the first Ca^{2+} peak in the stigma, the pollen itself displays a Ca^{2+} peak at the PT germination site. This suggests pollen uptake of external stigmatic Ca^{2+} and indicates that Ca^{2+} serves as an intracellular signaling molecule triggering PT germination.

2.2.2 Pollen Tube Guidance in the Transmitting Tract

In the transmitting tract (TT), the PT is exposed to female guidance cues, nutrients and growth promoting factors secreted to the extracellular matrix of the TT cells. Among the stylar factors are cell wall macromolecules like pectins and AGPs, as well as small secreted cysteine-rich proteins (CRPs), gamma-aminobutyric acid (GABA) and aminoacids, glycolipids, sugar and lipids.

PT growth in the stigma and style depends on the adherence of the PT to the extracellular matrix of the TT cells. In the hollow style of Lily, this is mediated by the concerted action of a stylar pectin and a cysteine-rich protein called STIGMA/STYLAR CYSTEINE-RICH ADHESIN (SCA), which have been purified from the stylar exudate and are able to promote PT adhesion to an artificial stylar matrix *in vitro* (*78*, *79*). In addition, SCA has been shown to bind to the tip of growing PTs *in vitro* (*80*). Taken together, these results suggest that SCA might act as a lectin on the PT surface coupling the PT to the style by interaction of SCA with the stylar pectin (*81*). This promotes adhesive PT growth along the style, which is referred to as haptotactic PT guidance. In *Arabidopsis*, the SCA-like lipid transfer protein LTP5 is present in the TT as well as in pollen, and although *ltp5* pistils only show minor defects, the gain-of-function mutant mainly affects the pollen side, as mutant PT growth is impaired both in wild-type pistils and *in vitro* experiments (*82*). Additionally, other small secreted molecules including *A. thaliana* plantacyanin and *Lily* chemocyanin are involved in PT guidance (*83*, *84*). Both peptides belong to the class of phytocyanins, an ancient blue copper protein family of unknown function localized in the extracellular matrix of the TT where they are regulating oriented PT growth in concert with SCA by acting as chemotropic signals through an unknown mechanism. In tomato, the low molecular weight polypeptide *Lycopersicum esculentulm Le*STIL (STYLE INTERACTOR FOR *Le*PRKs) has been identified as the female-derived ligand that binds to pollen-expressed receptor kinases (PRKs), inducing their dephosphorylation and thus positively regulating PT growth (*85*, *86*). Another pistil-derived ligand of *Le*PRKs, *Le*STIG1 was found to positively regulate PT growth (*87*). It is conceivable that species-specific binding of these receptor-ligand pairs could be involved in the recognition of PTs in the style.

The AGP TRANSMITTING-TRACT SPECIFIC (TTS) has been identified in the extracellular matrix of Tobacco styles as a factor promoting PT growth and serving as PT attractants *in vitro* (*88*). Stylar TTS proteins can be integrated into the PT cell wall where they are deglycosylated, suggesting that the sugar residues are used as nutrients by the PT (*89*). Interestingly, TTS proteins display a gradient of increasing glycosylation levels from the stigmatic to the ovarian end of the style, indicating that the sugar moieties on the proteins are the guidance cues for the PT. It is imaginable that TTS or related proteins confer species-specific recognition of PTs, as TTS has been implicated in the recognition of self/non-self pollen in the *S*-RNase based self-incompatibility system (see below,

(*90*)). Interspecific PT rejection (for example in *Nicotiana*) occurs frequently at the level of PT growth in the pistil and taken together, this raises the possibility that TTS-mediated nutrient supply and guidance are controlling PT rejection in the pistil (*35*, *38*).

In *Arabidopsis*, the glycosylphosphatidylinositol (GPI)-anchored protein COBRA-LIKE10 (COBL10) has been shown to regulate the PT's response to female guidance cues. It is localized to the apical plasma membrane of growing PTs and thus constitutes a likely receptor for female guidance cues (*91*). In *cobl10* mutants, PT growth is slower compared to wild-type PTs in the style, but not *in vitro*, and additional defects in micropylar PT guidance were observed.

2.2.3 – Funicular Guidance in the Ovary

Once the PT has reached the ovary, it needs to navigate towards the septum surface, exit the TT and grow along the funiculus towards the micropylar opening of the ovule. Guidance cues in these stages of fertilization are derived from the ovules and embryo sacs and are referred to as funicular guidance (leading to PT exit of the TT, both sporophytically and gametophytically controlled in *Arabidopsis*) and micropylar guidance (leading to PT growth towards the micropyle, in *Arabidopsis* solely relying on female gametophytic guidance cues), respectively (*92*).

One agent that guides the PT to exit the TT and emerge on the septum surface is the signal molecule GABA. GABA forms a concentration gradient along the whole pollen tube path with low levels at the stigma and highest levels at the septum surface cells and the ovule integuments near the micropyle (*93*). In addition, plants producing an excess of GABA in the pistil, such as mutants in the GABA-degrading transaminase *A. thaliana pollen-pistil 2 (pop2)*, display aberrant PT guidance. The GABA receptor has not been identified yet, but it could be shown in tobacco that GABA binds to cell-membranes of pollen protoplasts and regulates downstream Ca^{2+} oscillation, indicating that pistil-produced GABA is stimulating PT growth by influencing its internal Ca^{2+} levels (*94*). GABA is produced in the sporophytic ovule integument cells and the high GABA concentration at the micropylar opening might not only function in establishing the pistil gradient, but also in slowing down the PT at arrival, since it was shown *in vitro* that PT growth is inhibited at high GABA concentrations (*93*). Furthermore, the rare amino acid D-Serine and nitric oxide (NO) seem to be involved in funicular PT guidance (*95*,

96). Similarly to GABA, both compounds are produced in sporophytic ovule cells and mediate PT attraction by modulating cytosolic Ca^{2+} levels in the PT. The importance of the integument cells in PT guidance is further strengthened by the fact that among other female gametophyte developmental mutants, *A. thaliana inner-no-outer mutants* (*ino*), although displaying a functional female gametophyte, are lacking the outer ovule integument layer and are defective in funicular guidance of wild-type PTs (*97*).

On the male side, only a few genes have been identified that are involved in female signal perception of the PT. Among them are glutamate receptor-like channels (GLRs), that were shown to bind D-Serine secreted from the ovule integuments and subsequently form Ca^{2+} channels in the PT membrane, leading to an increased Ca^{2+} concentration at the tip (*95*). In addition, a pair of pollen-expressed cation/H^+ exchangers (*CHX21* and *CHX23*) has been implicated in perceiving ovular guidance cues because double mutant pollen tubes failed to re-orient their growth in order to leave the TT (*98*). It is believed that CHX21 and CHX23 are activated by so far unknown molecules and establish a new polarity axis by altering the pH gradient. Recently, a pair of mitogen-activated protein kinases, MPK3 and MPK6, have been identified to mediate PT response to female guidance cues and are proposed to act downstream of receptor kinases by direct binding to female signals (*99*). Whereas *mpk3 mpk6* double mutant pollen tubes showed impaired funicular guidance, their response to micropylar cues in a semi *in vitro* assay was normal, further indicating that funicular and micropylar guidance act via distinct female signals activating different signaling pathways in the PT. Further insight on the reaction of the PT to female guidance cues comes from a study that identified transcripts highly abundant in *in vivo* grown PTs, and mutants in the respective genes show defects in PT guidance (*iv6*, defective in a Xyloglucan endotransglucosylase/hydrolase gene), PT burst (*iv2*, a mutation in a methyl esterase) and the repulsion of multiple PTs (a mutant in a Glutathione S-transferase gene, *iv4*, as well as *iv2*) (*70*).

In interspecific crosses between *A. thaliana* females and *Brassica oleracea* pollen, the pollen germinate and PTs grow through the style, but cannot exit the TT in order to reach the ovules, indicating that in this particular case, intra- and interspecific signals differ only at the stage of funicular guidance (*100*). In contrast, if the reciprocal cross is conducted, the interspecific barrier already acts similar so the self-incompatibility response at the stage of *A. thaliana* pollen germination on *B. oleracea* stigma.

Unfortunately, the molecular basis of interspecific incompatibility at the level of funicular guidance remains unknown.

2.2.4 – Micropylar Guidance

Short-range micropylar guidance signals are derived from the cells of the female gametophyte itself, more specifically it is mainly the synergid cells that are responsible for secreting these guidance cues (*101*). Most of the knowledge on micropylar guidance is derived from studies in *A. thaliana* and many genes regulating female gametophyte development have been found to be affecting PT guidance (*92*). This is not surprising as it was found that PT attractants, such as *At*LURE1s (see below), are only secreted when the gametophyte is fully mature (*37*). Because some of the gametophytic mutants are mainly affecting general house-keeping genes and are thought to only have an indirect effect on signaling during PT guidance, they will not be further discussed in this introduction. Among them are the SUMO E3 ubiquitin ligase *SIZ2-1*, the nuclear protein CENTRAL CELL GUIDANCE (CCG), *MAGATAMA1 (MAA1)* and the helicase *MAA3*, the disulfide-isomerase *PDIL2-1*, and the transcription factor *MYB98* (*102-106*). However, it is noteworthy to point out that *CCG* is a central-cell specific gene, indicating that in *Arabidopsis* not only the synergids but also the central cell play a critical role in PT attraction (*103*). Also, the egg cell seems to be involved in micropylar PT guidance, at least in *Arabidopsis* and maize (*107*, *108*). Besides these genes that seem to play indirect roles in micropylar PT guidance, several genes directly involved in the cross-talk between male and female gametophytes have been identified (see below). Interestingly, many of them have been implicated to act in a species-preferential manner.

Initial laser ablation studies in *Torenia founieri*, which possesses protruding embryo sacs facilitating such experiments, showed that it is the synergid cells that are secreting PT attractants (*101*). After fertilization of the embryo sac, the secretion of the attractants ceases, thus preventing multiple PTs from being guided towards the female gametophytes. Using five different *Torenia* species as pollen or embryo sac donors, it could be shown that the synergid-dependent attraction of PTs is species-preferential (*109*). The guidance cues emitted by the synergids turned out to be low molecular weight defensing-like cysteine-rich proteins (*T. fournieri CRP*s or *LURE*s), which could attract PTs species-specifically *in vitro* (*36*). It could be shown that LUREs directly bind to the tip of PTs that have grown through a style in a semi-*in vitro* assay (*110*). Because

the ability of PTs to bind LUREs was strongly reduced if they had not grown through a style and the PT transcriptome was changed after passing the style, the authors concluded that earlier male-female interactions are crucial to prepare the PT to be susceptible to LURE signaling. In *Arabidopsis*, six defensin-like peptides (*At*LURE1.1 – 1.6) attract *A. thaliana* PTs with higher efficiency than PTs from *A. lyrata*, whereas their *A. lyrata* homologs (*Al*LURE1s) do not display such a species-specificity (*37*). Heterologous expression of *AtLURE1* in *T. fournieri* resulted in strong attraction of *A. thaliana* PTs and even egg apparatus penetration by the PT was observed, an effect implying that *AtLURE1s* can overcome prezygotic hybridization barriers despite the large evolutionary distance between *Arabidopsis* and *Torenia* (*27*). However, fertilization of the egg cell failed in these experiments. The expression of *AtLURE1*s is dependent on the synergid-specific transcription factor *MYB98*, hence in *myb98* mutants, *At*LURE1s are down-regulated (*111*). In maize, *ZmEA1* (*Zea mays EGG APPARATUS1*) was identified as an egg and synergid-cell specific, small secreted hydrophobic protein that is critical for short-range PT guidance at the micropyle (*108*). Indeed, chemically synthesized *Zm*EA1 binds to the surface of PTs in a species-specific manner and is rapidly internalized and degraded (*112*, *113*). Similar to *At*LUREs1, *Zm*EA1 could also be used to overcome hybridization barriers since transgenic *A. thaliana* ovules expressing *ZmEA1* in their egg apparatus were able to attract maize PTs *in vitro* (*112*). Both LUREs and *Zm*EA1 are likely to be ligands for PT-localized receptor kinases or ion channels, and binding of the ligands would trigger intercellular responses in the PT leading to oriented growth towards the micropylar opening of the ovules. Indeed, recently a pair of PT-specific, plasma membrane-localized, receptor-like cytoplasmic kinases, *LOST IN PT GUIDANCE1* (*LIP1*) and *LIP2*, has been identified (*114*). Interestingly, *lip1 lip2* double mutant pollen tubes are not targeted to the micropyle *in vivo* and show reduced attraction to synthetic *At*LURE1 *in vitro*, indicating that LIP1 and LIP2 might be the pollen receptors that directly interact with female guidance attractants. Another determinant for the PT response to female cues is POD1 (named after its mutant phenotype, *pollen defective in guidance 1*), which is a protein present in the ER lumen and therefore does not seem to be directly involved in the crosstalk between PT and embryo sac (*115*). Its effect on PT guidance might be due to its interaction with the Ca^{2+}-binding chaperone CALRETICULIN3 (CRT3), and thus, POD1 might work in concert with CRT3 in ER quality control of putative transmembrane receptors that perceive signals of the female gametophyte.

2.3 – Pollen Tube Reception

After the PT has successfully arrived at the embryo sac, PT reception begins with PT growth arrest, programmed cell death of the receptive synergid cell and PT burst, followed by the release of the two sperm cells into the receptive synergid, and finally gamete fusion (discussed in section 2.4). Extensive communication is needed between the male and the female gametophytes, as at this stage it is one of the last chances to discriminate species before fertilization. Similar to micropylar PT guidance, signaling during PT reception is mostly conducted by the synergid cells (Fig. 1.2). These cells at the micropylar end of the female gametophyte possess a thickened cell wall structure, which is membrane-rich and called the filiform apparatus (FA). It is believed that this structure accounts for most of the communication events during PT guidance and reception (*53*). However, it has been shown that PT penetration of the receptive synergid does not occur at the site of the FA, but rather at a more distant zone called the "synergid hooks" characterized by cytoplasmic protrusions of the central cell (*116*). It is conceivable and consistent with observations of (*117*), that upon arrival at the FA, the PT slows down its growth and gets primed by communication with the synergids, and only then continues its growth along the synergid to reach the hook where the actual penetration of the receptive synergid occurs.

Cell-cell communication between synergids and PT is to a great extent mediated by coordinated Ca^{2+} oscillations both in the PT and in the receptive and non-receptive synergid (*117*). In contrast, in *feronia* (*fer*) mutant ovules which are impaired in PT reception, this Ca^{2+} dialogue is disturbed and shows a distinct pattern to the one observed in wild-type synergids and PTs, indicating that Ca^{2+}-mediated crosstalk is not only crucial for PT germination, growth and guidance, but also for PT reception. In *fer*, which is allelic to *sirène* (*sir*), the defect in PT reception is characterized by a failure of communication between the gametophytes, leading to continuous PT overgrowth within the embryo sac (*118*, *119*). Mutant ovules with PT overgrowth remain unfertilized even by wild-type PTs, because the PT does not rupture to release its sperm cells. *FER* codes for a *Cr*RLK1L (*Catharanthus rhoseus* RLK1-like) receptor-like serine-threonine kinase, a gene family comprising 17 members in *A. thaliana* which is further described in a separate introductory chapter of this thesis (*120*).

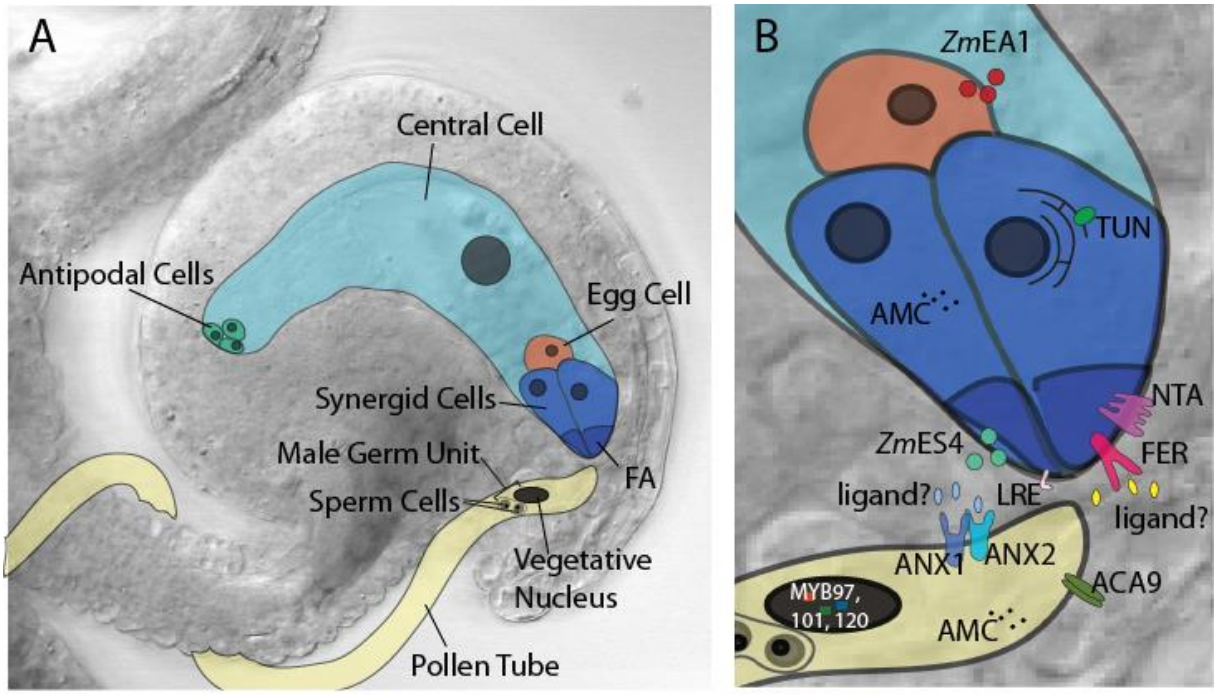


Figure 1.2: Cell types of the female gametophyte and molecular factors involved in pollen tube reception. (**A**) The female gametophyte consists of four cell types. The synergids possess a specialized structure at their micropylar end, the filiform apparatus (FA). Sperm and vegetative nucleus travel together in the PT in a male germ unit. (**B**) Proteins involved in PT reception. FERONIA (FER), a receptor-like kinase and its putative ligands; NORTIA (NTA), a MLO protein; LORELEI (LRE), a GPI-anchored protein; the maize secreted proteins *Zm*ES4 and *Zm*EA1; the UDP-glycosyltransferase TURAN (TUN) at the ER; ABSTINENCE BY MUTUAL CONSENT (AMC) at the peroxisomes; three MYB transcription factors (MYB97, 101, 120); the Ca^{2+} ATPase ACA9 and the receptor-like kinases ANXUR 1 and 2 (ANX1, 2) with their unknown ligands.

It is believed that the membrane-localized FER, which contains an extracellular carbohydrate-binding malectin-like domain, perceives putative signals from the arriving PT (that could be either directly derived from the PT or indirectly caused by it during penetration of the embryo sac cell walls) (*121*, *122*). Subsequently, this would trigger intracellular responses in the synergids, which would in turn lead to feedback signaling to the PT and finally cause PT rupture. Such putative signals from the PT could be species-specific and indeed, it has been observed that interspecific crosses between different species within the Brassicaceae and the Ericaceae families display a PT overgrowth phenotype reminiscent of *fer* (*121*, *123*, *124*). Recently, the secreted rapid-alkalization factor peptide RALF has been identified as the ligand of FER in roots, where the receptor-kinase functions in controlling cell elongation (*125*). In contrast, the ligand of FER in synergids remains elusive. On the female side, other signaling components in the PT reception pathway have been identified and some of them are membrane-

localized, making them potential candidates for direct interactions with the PT: the glycosylphosphatidylinositol (GPI)-anchored protein LORELEI (LRE), the Mildew Resistance Locus O (MLO)-like protein NORTIA (NTA), and the UDP-Gylcosyltransferase TURAN (TUN) (*126-128*). Similar to *fer*, *lre*, *nta* and *tun* mutants fail to receive the PT and display PT overgrowth. In addition, *scylla* (*syl*), another female gametophytic mutant with PT overgrowth has been described, but the underlying gene has not been mapped (*129*). LRE is a synergid surface protein and might act directly in the interface of the female gametophyte and the PT, either as a binding partner for PT signals or as a secreted signal from the embryo sac, and possibly in concert with FER (*126*). While the connection between FER and LRE, remains speculative, it has been shown that the membrane protein NTA becomes polarly localized towards the FA in a *FER*-dependent manner after PT arrival (*127*). The fact that a protein involved in the protein N-glycosylation pathway (TUN) plays a role during PT reception suggests that it could be involved in the glycosylation of ligands or receptors, which might be crucial for their function or binding (*128*). A candidate target for TUN-mediated glycosylation might be FER, as it has eight predicted glycosylation sites in its extracellular domain (*120*), however, experimental validation is missing. Moreover, the peroxin gene *ABSTINENCE BY MUTUAL CONSENT* (*AMC*) has been shown to be essential for PT reception and to regulate PT discharge (*130*). However in *amc*, unlike in the above-described female gametophytic mutants, the PT overgrowth phenotype is only observed when both the PT and the embryo sac are lacking functional AMC at their peroxisomes. These findings indicate that diffusible compounds produced in the peroxisomes, such as nitric oxide (NO) or reactive oxygen species (ROS), are involved in the dialog leading to PT reception (*130*). Supporting this, ROS signaling has been implied in the *FER*-pathway during tip growth of root hairs, where FER is upstream of RAC/ROP GTPases which subsequently activate ROS-producing NADPH oxidases, finally leading to extracellular ROS accumulation, cell wall loosening and cell expansion (*131*). The downstream signaling components of FER in synergids are not known, however a recent study demonstrates that ROS accumulates at the FA in a *FER*- and *LRE*-dependent manner and is essential for PT burst, as it was shown in a pistil feeding assay that inhibition of ROS production led to PT overgrowth and that addition of ROS to semi-*in vitro* growth PTs induced burst by increasing the levels of PT cytoplasmic Ca^{2+} (*132*). In maize, the female gametophyte-specific, cysteine-rich defensin-like protein *Zm*ES4 (*Zea mays* EMBRYO SAC4) induces *in vitro* PT growth arrest and burst in a species-preferential manner by opening a

potassium channel in the PT inducing plasma membrane depolarization (*133*). *Zm*ES4 protein accumulates in the secretory zone of the synergids, and in *ZmES4*-RNAi-lines, PT burst is impaired and PT overgrowth is observed. Pectin methylesterases (PME) are crucial components for mediating cell wall integrity and stability and are tightly regulated by PME inhibitors (PMEIs) (*134*). In maize, the male and female gametophytic expressed *Zm*PMEI1 induces PT burst, but not growth arrest, when applied to *in vitro* growing PTs and it is believed that *Zm*PMI1 is secreted by the female gametophytic cells in order to destabilize the PT cell wall (*135*). In addition to its function in micropylar pollen tube attraction, *Zm*EA1 also seems to have a function in pollen tube growth arrest when applied in high concentrations (*108*). Taken together, these findings indicate that PT growth arrest and burst are regulated by different pathways.

Whereas several female factors involved in PT reception and burst have been identified, the players on the pollen side remain largely unknown. Presumably the best-studied genes are the *Cr*RLK1L receptor-like kinases and closest homologs to *FER*, *ANXUR1* (*ANX1*) and *ANX2*. ANX1 and ANX2 act together in controlling cell wall integrity and stability of the PT by activating a signaling cascade leading to NADPH-oxidase-dependent ROS production and furthermore to the establishment of a tip-focused Ca^{2+} gradient (*136*). Consequently, double mutant *anx1 anx2* PTs burst shortly after germination and overexpression of ANX1 induces PT growth arrest and an overaccumulation of cell wall material (*136*-*138*). Because ANX1 and ANX2 have a similar carbohydrate-binding domain like FER, and they seem to have opposite functions (FER promotes PT burst, ANX1/2 prevent it), it was proposed that the receptors compete for the same ligand when PT and embryo sac meet, leading to the inactivation of the ANX1/2 pathway in the PT and the activation of the FER pathway in the synergids (*139*). Another pollen expressed gene involved in sperm cell discharge is the plasma membrane-localized autoinhibited Ca^{2+} ATPase *ACA9* (*140*). Mutants in *aca9* have PT growth defects and partially fail to burst upon arrival at the synergids. Although details about interactions with potential female signals remain elusive, the importance of Ca^{2+} signaling during PT reception is apparent. Only recently, the first male gametophytic mutant with *fer*-like overgrowth phenotype has been isolated: *myb97*/*myb101*/*myb120* triple mutant PTs fail to arrest their growth and to discharge their sperm cells in wild-type ovules (*141*, *142*). These MYB transcription factors regulate the expression of a variety of genes, including transmembrane transporters,

carbohydrate-active enzymes, and small secreted proteins (*141*), and it is conceivable that some of these mediate direct interaction with components of the female gametophyte. Only a short time ago, it was shown that the position of the vegetative nucleus within the PT is crucial for PT reception success. During wild-type PT reception, the vegetative nucleus precedes the two sperm cells in the male germ unit, possibly because transcripts from the vegetative nucleus might be required for the cross-talk with the synergids at the PT tip (*143*). In mutants of two outer membrane proteins that interact with myosin (*wip*, deficient in *WPP DOMAIN-INTERACTING PROTEINS*; and *wit*, lacking *WPP DOMAIN-INTERACTING TAIL ANCHORED PROTEINS*), the male germ unit is disorganized, with the sperm cells travelling ahead. This leads to defective PT reception of mutant PTs in more than 50% of the ovules, characterized by stalled PTs that don't burst or PT overgrowth (*143*).

2.4 – Gamete Fusion

After PT burst in the receptive synergid, the released sperm cells fertilize egg and central cell. In order to achieve this so-called double fertilization, gamete recognition and fusion need to be tightly controlled. After successful gamete recognition and sperm activation, the plasma membranes of the gametes fuse (plasmogamy), the sperm content enters the egg cell, and subsequently also the nuclei fuse (karyogamy) (*144*).

Initial gamete adhesion and recognition is controlled by the sperm surface protein *Arabidopsis* GAMETE EXPRESSED 2 (GEX2), which is rapidly evolving and therefore constitutes a good candidate for mediating direct, species-specific gamete recognition (*145*). GEX2 possesses extracellular immunoglobulin-like domains, reminiscent of the protein structure of the mammalian sperm surface protein Izumo (*146*), and in ovules targeted by *gex2* mutant PTs, gamete attachment fails. Another key player during gamete interaction is the sperm membrane protein encoded by *GAMETE CELL SPECIFIC* (*GCS*), which is allelic to *HAP2*. In *gcs/hap2* mutants, the entire fertilization process including sperm cell release is functional, but instead of fusing with the female gametes, the sperm cell pair remains at the site of the degenerated receptive synergid cell (*147*, *148*). Supporting the GCS/HAP2 role as a fusogen, species-specific gamete fusion in *Chlamydomonas* and *Plasmodium* is *GCS/HAP2*-dependent, further highlighting a conserved role of the gene across species (*149*). In sperm cells, GCS/HAP2 is located in

the endomembrane system, but not on the surface as it would be expected for a direct interactor of egg surface proteins (*150*). Yet, it has been shown that GCS/HAP2 is transported to the sperm cell surface after exogenous application of recombinant *Arabidopsis* EGG CELL 1 (EC1) (*150*). This process is called "sperm activation" since the sperm cell exposes its fusogenic regions, making it able to connect with the egg cell. *EC1* belongs to a gene cluster comprising five genes (*EC1.1* – *EC1.5*), which are encoding small, cysteine-rich proteins that are female gametophyte specific (*150*). Before fertilization, EC1.1 is localized in vesicles in the egg cell, and gets secreted into the degenerating receptive synergid after the PT has released the sperm cells, thus forming a gradient possible guiding the sperm cells towards the egg. Similar to the egg cell, the central cell also promotes its own fertilization (*151*). In mutants of the central cell-expressed acyl-transferase *glauce* (*glc*), only the egg but not the central cell gets fertilized, resulting in an embryo which can develop up to the globular stage, but no endosperm development takes place. The mitochondrial protein ANKYRIN6 (ANK6) has been implicated in gamete recognition being essential in both female and male gametes, and it is mediating this process possibly by regulating mitochondrial gene expression together with the σ-factor SIG5 (*152*). Both GLC and ANK6 are most likely only indirectly regulating plasmogamy by acting as modifiers of other molecular factors directly involved in gamete fusion. To date, no genes important for karyogamy have been identified and although there are many opportunities for species barriers during gamete recognition, experimental proof is missing.

After successful gamete fusion, additional arriving PTs have to be repelled in order to prevent fertilization of the female gametes by more than one sperm (polyspermy). Although there is evidence that polyspermy barriers exist in plants, the genetic basis of their establishment is unknown (*153*). The attraction of multiple PTs (polytubey) has been observed in the *fer*, *nta*, *lre* and *amc* PT reception mutants, in ovules pollinated with *gcs*/*hap2* pollen as well as in *ec1* mutants, but the underlying mechanism remains elusive (*118*, *126*, *127*, *130*, *150*, *154*). However, it implies that if double fertilization has not occurred, additional PTs are attracted to accomplish sperm delivery, a phenomenon improving fertilization success termed fertilization recovery (*154*, *155*). Interestingly, both gametes have to be fertilized in order to initiate a polytubey block, indicating that both egg and central cells possess their own signaling pathway (*156*). Recently, it could be shown that the second non-receptive synergid undergoes ethylene signaling-

dependent programmed cell death after gamete fusion, thus preventing it from emitting further guidance cues and terminating the fertilization process (*157*). It is conceivable that besides the decrease of PT guidance signaling after synergid degeneration, the embryo sac emits additional, repulsive signals to prevent polytubey (*92*). However, there are no studies yet to prove this assumption.

2.5 – Linking Self-Incompatibility to Interspecific Incompatibility

The above described molecular factors implied in the recognition and cross-talk of male and female tissues could potentially be acting species-discriminant. One particular case of pollen-discrimination is self-incompatibility (SI). In contrast so interspecies incompatibility systems that prevent cross-pollination, SI systems promote outcrossing, however the molecular mechanisms between both systems could be highly similar. More than 50% of all Angiosperms are SI (*92*), and the known SI systems can be grouped by their functional mechanisms: i) the sporophytic SI of Brassicaceae, ii) the gametophytic, *S*-RNase based SI of Solanaceae, Rosaceae and Scrophulariaceae, and iii) the gametophytic SI of Papaveraceae. The SI systems found in Brassicaceae and Papaveraceae (i and iii) are based on self-recognition that depends on specific interactions of the male and female determinant originating from the same *S* haplotype. In contrast, *S*-RNase dependent SI systems (ii) constitute a non-self recognition system. In this case, the recognition occurs between male and female determinants from different *S* haplotypes (*158*).

i) The Brassicaceae SI system has been investigated in great detail using *Brassica* spec and *A. lyrata*. The highly polymorphic, multigenic *S*-locus encodes for both the male (*S-LOCUS CYSTEINE-RICH PROTEIN, SCR*; or *S-LOCUS PROTEIN, SP11*) and the female determinant (*S-LOCUS RECEPTOR KINASE, SRK*), whose gene products have been shown to interact directly with each other if they are derived from the same haplotype (*159-162*). In *Brassica* but not in *A. lyrata*, the *S*-locus additionally codes for the *S*-LOCUS GLYCOPROTEIN (SLG), a secreted stigma protein enhancing SI response (*161*, *163*). *SCR* is expressed in tapetum cells of anthers and the protein is transferred to the pollen coat during pollen development (*164*). The female factor SRK is localized to the plasma membrane of (sporophytic) papilla cells where the SI reaction takes place, which is why this SI system is considered to be sporophytic (*159*). During self-pollination, SCR binds

to the respective SRK and stabilizes and activates it in a dimeric form at the plasma membrane. The activation of SRK by SCR promotes the recruitment and subsequent phosphorylation of the M-LOCUS PROTEIN KINASE (MLPK) which then activates the E3 ubiquitin ligase ARM-REPEAT CONTAINING 1 (ARC1), a putative negative regulator of the exocyst complex subunit EXO70A1. These reactions lead to degradation of the SRK/SCR complex as well as EXO70A1, and finally to pollen rejection (*165*). In compatible (non-self) pollinations, SRK is not activated by SCR and thus, pollen hydration involving EXO70A1 is initiated (see section 2.1; (*63*)). In *A. thaliana*, SI was lost due to loss-of-function of *SCR*, whereas the female determinant is still active in some accessions like Wei-1. Restoration of *SCR* in Wei-1 lead to pollen rejection on the stigma of an otherwise self-compatible species (*166*).

Whereas in the *Brassica* SI system pollen rejection occurs at the hydration step on the stigma, the *S*-RNase based SI system (ii) acts during PT growth in the transmitting tract. Here, the female determinant is the *S*-locus-encoded stigma- and style-secreted *S*-RNase, whereas the male determinant is an F-Box protein (*S-LOCUS F-BOX, SLF*) (*167*). In addition, other molecular factors not encoded by the *S*-locus are essential for SI, such as HT-B, a small asparagine-rich protein, and the stylar glycoproteins TRANSMITTING TRACT SPECIFIC (TTS), PISTIL EXTENSIN-LIKE PROTEIN III (PELP III) and 120K (*90*, *168*). The latter three have been shown to directly bind *S*-RNase *in vitro* and it is hypothesized that in the style, this complex forms a recognition interface with the pollen tube (*169*). Upon compatible PT arrival in the transmitting tract, the *S*-RNase enters the PT and a model has been proposed in which the *S*-RNase directly binds SLF derived from a non-self *S* haplotype (*158*, *170*). In this case, the *S*-RNase is degraded and PT growth continues. If SLF originates from the same *S* haplotype like the *S*-RNase (self-pollination), SLF and *S*-RNase do not interact and thus, the *S*-RNase acts as a cytotoxic agent degrading PT RNA and hence inhibiting PT growth.

Similar to the *Brassica* SI system, the SI system in *Papaver* (iii) is based on the recognition of male and female determinants comprising the same *S* haplotype. Like in *S*-RNase based SI, the interaction of the SI factors occurs in the male gametophyte, which is why these two types of SI are referred to as gametophytic SI. In *P. rhoeas*, the female determinant is a stigma-secreted, small protein called *Pr*sS (*P. rhoeas* stile *S*), and the pollen determinant, *Pr*pS, is a plasma-membrane localized protein that directly interacts with *Pr*sS (*171*). Upon self-pollination, *Pr*sS binds to *Pr*pS, a recognition process which

subsequently triggers PT growth arrest and programmed cell death in incompatible (self) pollen (*172*). The downstream SI response is well characterized and involves the activation of a mitogen associated kinase (MAPK) as well as the increase of intracellular levels of Ca^{2+}, ROS, NO, leading to actin depolymerisation and finally to the activation of the programmed cell death signaling cascade (*173*).

SI species are usually more selective in interspecific crosses than self-compatible (SC) ones, which are more likely to accept pollen from an SI species than vice versa (*174*). This phenomenon is called unilateral incompatibility (UI), and some but not all loci involved in UI have been mapped to be tightly linked to the *S*-locus, thus providing a hint that SI and interspecific incompatibilities might be at least partially controlled by similar mechanisms (*35*, *175-178*). In a recent publication, Tovar-Mendéz et al. showed that combined expression of *S-RNase* and *HT-B* from multiple SI green-fruited tomato species in the normally SC red-fruited tomato *Solanum lycopersicum* could recapitulate both SI and UI (*179*).

The crosstalk between pollen and female tissues is essential for recognition of incompatible pollen. SI and UI share common mechanisms, but to date, much more is known about the molecular basis of SI than of UI. Because interspecific incompatibilities act at the species level, they are less specific than SI barriers, which have to distinguish between self- and non-self pollen of a single species. Moreover, interspecific barriers seem to be more complex than SI because they usually combine multiple mechanisms at various levels during the reproductive process (*3*).

3 – The Role of Glycosylation in Cell-Cell Interactions

Cell-cell communication processes not only play a critical role during fertilization, but are also crucial for development (where cells from the same individual communicate with each other) as well as in immune response and symbiosis (where cells derived from different organisms communicate). Often, this cross talk relies on secreted ligands and their corresponding receptors and some examples of known receptors-ligand pairs in reproduction have been discussed in section 2. All proteins in the secretory pathway, including cell surface receptors and membrane-bound or secreted ligands, are subject to ER quality control, a mechanism that ensures that only proteins with correct folding and

correct modifications, e.g. glycosylation, are secreted to the plasma membrane (*180*). N-glycosylation affects a protein's stability, folding, targeting and binding specificity (*181*).

Many proteins involved in various steps of fertilization are heavily glycosylated, for example the *S*-locus factors SRK and SLG mediating *Brassica* self-incompatibility, the AGPs TTS involved in PT guidance and possibly also the GPI-anchored proteins COBL10 and LRE, which regulate PT guidance and reception, respectively (see section 2 for details). In addition, the receptor-like kinases FER and ANX1/2 contain several predicted glycosylation sites in their extracellular domain (*120*). Likewise, interaction and recognition of mammalian gametes also relies on the action of heavily glycosylated receptors and their ligands. During development, the mammalian egg cell secretes an extracellular matrix called zona pellucida (ZP), which consists of various ZP glycoproteins (in mouse: ZP1, ZP2, ZP3) directly involved in sperm binding (reviewed in (*182*)). Sperm binds to the ZP in the so-called "acrosome reaction", which leads to exocytosis of the contents of a secretory vesicle (acrosome) in the sperm's head, and subsequently to sperm penetration of the ZP. Initial sperm-binding to ZP3 is species-restricted (*183*). Whether the glycosylation status of ZP3 is critical for this process is debated, because *in vitro* studies abolishing specific glycosyl-residues on ZP3 show reduced sperm-binding, whereas *in vivo* experiments with mice completely lacking these residues did not show a fertility defect (reviewed in (*184*)). These contradictory results lead to the proposition of the "domain-specific model", in which a sperm protein interacts both with the carbohydrate residues and the protein backbone of ZP3 (*184*). In contrast to the somewhat unclear situation in mouse, sperm-oocyte interaction in abalone relies on lectin-like, species-specific binding of the sperm protein lysin to the glycosylated egg receptor VERL (Vitelline Envelope Receptor for lysin) (*185*).

3.1 – Signaling during Plant-Microbe Interaction

Plant defense and fertilization share common molecular components, and also during pathogen-host communication, species-specific signaling events take place allowing only certain pathogen species to penetrate a plant cell. The connection between defense and PT reception has been discovered for *FER* and *NTA*, which both play a critical role in PT reception (see section 2.3). *NTA* is a member of the Mildew Locus O protein family, which has originally been implicated in mediating resistance to powdery mildew (*186*).

Powdery mildew hypha are, similar to PTs, tip-growing and during infection, MLO proteins are transported to the site of fungal invasion, similar to NTA being re-located to the filiform apparatus after PT arrival at the embryo sac (*127*, *187*). Supporting the connection between PT reception and plant defense, *fer* mutants are resistant to powdery mildew infections but not to other mycetes, indicating that FER is a putative compatibility factor and thus a critical player for (specific) fungal susceptibility (*127*).

Plant cells are able to sense wounding of their cell wall through so-called microbe- or damage-associated molecular patterns (MAMPs or DAMPs), which consist of specific molecules derived from the pathogen or can be fragments derived from the destruction of the plant cell wall (*188*). The plant cell senses these cues with the help of surface-localized pattern recognition receptors (PRRs) and reacts by complete rearrangements of its cytoskeleton and organelle movement, as well as by depositing callose at the wounding site, which is considered to be both a physical barrier due to a reinforced cell wall as well as a chemical barrier (*189*). Leucine-Rich Repeat receptors (LRR-PRRs) confer one important group of surface-exposed PRRs, and prominent examples are *FLAGELLIN-SENSITIVE 2* (*FLS2*) and *EF-TU RECEPTOR (EFR)* detecting the bacterial peptide effectors flagellin (flg22) and EF-Tu (elf18), respectively (*190*, *191*).

In contrast to LRR-PRRs that bind proteinaceous ligands, LysM-PRRs sense carbohydrates. Their Lysin Motif (LysM) is typically around 40-65 amino acids in length and its sequence is well conserved in the first 10-16 residues (*192*). The rest of the motif is variable with some highly conserved amino acids (Ile at position 20, Asn at 27 and Leu at position 30 of the consensus sequence of all available LysMs). The motif consists of a three-dimensional βααβ secondary structure and strongly binds to N-acetylglucosamine (GlcNAc), a structural component of fungal and bacterial cell walls (chitin and peptidoglycan, respectively) (*193*, *194*). Consequently, all so far described plant LysM proteins have been implicated in the recognition of MAMPs during plant defense and establishment of symbiosis (*195*). All known plant LysM proteins are plasma membrane localized containing one or more extracellular LysM domains (Fig. 1.3). Some of them possess an intracellular kinase domain, which is thought to play a role in the transduction of the signal. In addition, small peptides with only one LysM are believed to be either secreted or remain in the cytoplasm (*195*).

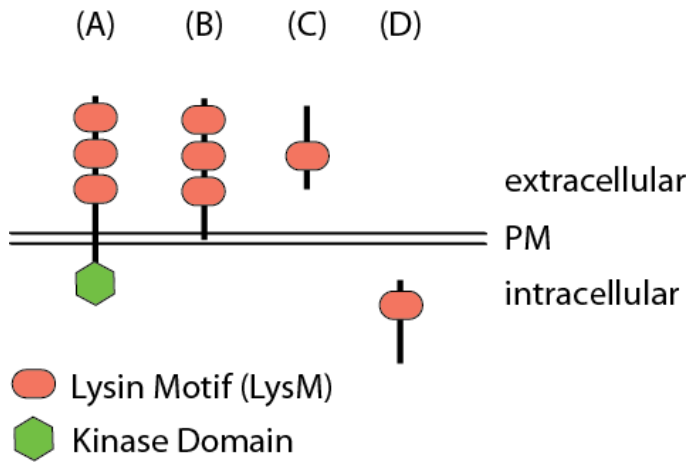


Figure 1.3: Plant LysM proteins and their localization. (**A**) Receptor-kinases with extracellular LysM domains and an intracellular kinase. Examples are CERK1, *Ps*SYM37, *Lj*NFR5 and *Mt*NFP. (**B**) membrane-anchored LysM proteins without a kinase domain such as CeBIP, LYM1 and LYM3. (**C**) small, secreted proteins with one LysM domain that are predicted to be secreted. (**D**) small peptides with only one LysM domain predicted to remain intracellular. For types (C) and (D), no examples are known to date. PM: Plasma membrane. Figure is adapted from (*195*).

Among the characterized plant LysM proteins are receptors involved in plant-pathogen cross-talk such as the CHITIN ELICITOR RECEPTOR KINASE 1 (CERK1) acting in concert with the outer membrane protein CEBiP (chitin elicitor binding protein) in rice (*196*), as well as *A. thaliana* CERK1, which unlike OsCERK1, does not require a co-receptor for chitin-induced signaling (*197*). AtCERK1 is also crucial for bacterial peptidoglycan binding, but in this case its function is dependent on the plasma membrane-tethered, LysM-containing proteins LYM1 and LYM3 (*198*).

During root nodule formation between nitrogen-fixing rhizobacteria and legumes, the bacteria secrete lipochitin-oligosaccharides called Nod-factors. These Nod-factors are recognized by LysM-domain containing receptor kinases such as the Nod-factor receptors NFR1 and NFR5 (*Lotus japonicus*), *Medicago truncatula* Nod-factor perception (MtNFP) and MtLYC3, as well as SYM37 and SYM10 from *Pisum sativum* (*199-202*). Finally, successful binding of bacterial Nod-factors to plant receptors will eventually lead to endocytosis of the bacteria and nodule formation. The only known receptor mediating arbuscular mycorrhiza formation between symbiontic fungi and the non-legume *Parasponium andersonii* is *PaNFP*, which is hologous to *MtNFP* and interestingly also mediates root nodule symbiosis in *P. andersonii* (*203*).

Although LysM receptors are heavily glycosylated, their glycosylation status does not influence ligand binding (*200*). Instead, LysM domains covalently bind their

carbohydrate substrate with critical amino acids forming a hydrophobic binding pocket. These amino acids determine substrate specificity of LysM. For example, mutation of a critical leucine in the LysM domain of *Lj*NFR5 abolishes its specific binding to *Rhozobium leguminosarum* Nod-factors but does not influence the response to *Mesorhizobium loti* (*204*).

3.3 – N-Glycosylation Mechanisms

N-glycosylation describes the co- and posttranslational transfer of a preassembled carbohydrate-oligomer ($Glc_3Man_9GlcNAc_2$) onto specific asparagine (N) residues within the glycosylation motif N-X-S/T (with X denoting any amino acid except proline) of a nascent protein in the ER lumen (Fig. 1.4). This process is catalyzed by the oligosaccharyltransferase (OST), a heteromeric protein complex consisting of nine subunits in yeast: Stt3p, Ost1p, Ost2p, Wbp1p, Swp1p, Ost3p, Ost4p, Ost5p and Ost6p, of which only the first five are essential for viability and Stt3p is the catalytical subunit that transfers the carbohydrate oligomer to the nascent polypeptide (*181*). The non-essential subunits play a role in substrate specificity and enhance the overall performance of OST. In yeast, the homologs Ost3p and Ost6p are present in two distinct OST isoforms and confer polypeptide substrate specificity (*205*, *206*). Glycosylation motives with the sequon N-X-T are more efficiently glycosylated than N-X-S sequons and in general, only about 60% of the glycosylation sites are actually glycosylated (*207*, *208*).

The assembly of the $Glc_3Man_9GlcNAc_2$ oligomer at the ER membrane is conducted by several enzymes both in the cytoplasm and in the ER lumen (reviewed in (*208*)). As a first step, the lipipd carrier dolichol situated in the ER membrane is phosphorylated by a dolichol kinase, which is the prerequisite for carbohydrate assembly. In yeast, the carbohydrate oligomer consists of three glucoses, nine mannoses and two GlcNAc residues, which are assembled subsequently by transferring them either from soluble, nucleotide-activated sugar donors or from dolichol-phosphate linked sugars. These processes are catalyzed by distinct groups of glycosyltransferases acting either in the cytoplasm (transfer of soluble sugars) or in the ER (transfer from dolichol-phosphate). The concerted action of the differentially localized glycosyltransferases requires flipping of the dolichol-bound glycan from the cytoplasmic to the luminal side of the ER membrane. After the carbohydrate oligomer assembly is complete, it is transferred en

bloc by the OST to the glycosylation site of the target protein and can be further processed in the Golgi apparatus in a tissue- and species-specific way (*209*).

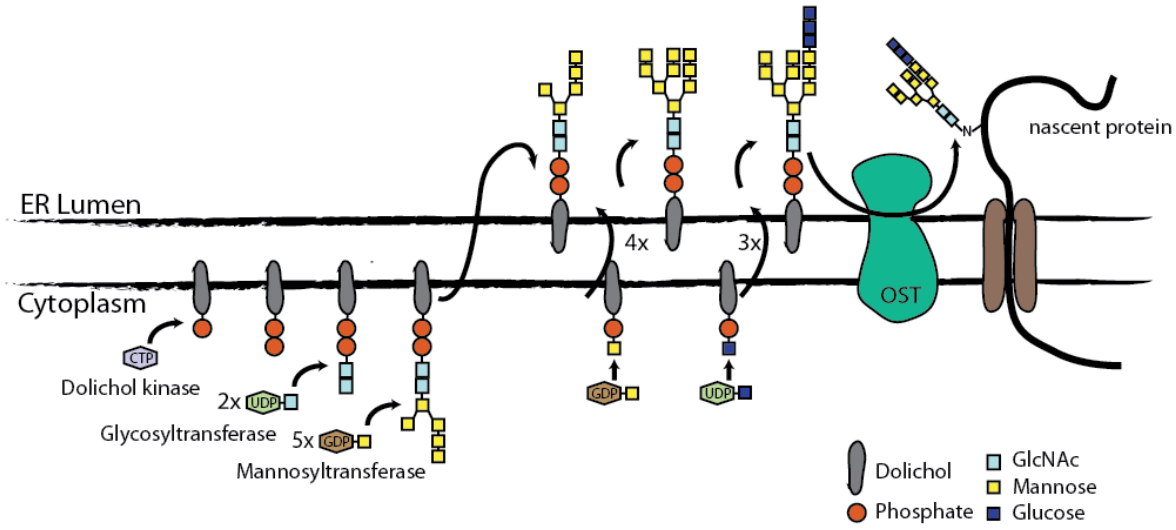


Figure 1.4: N-glycosylation at the ER membrane. Phosphorylation of a Dolichol-carrier by the Dolichol-Kinase, followed by attachment of N-acetylglucosamine (GlcNAc) and mannose residues. After flipping to the ER luminal side, the $Glc_3Man_9GlcNAc_2$ oligomer is completed by adding mannose and glucose molecules. Subsequently, the oligosaccharyltransferase (OST) transfers the carbohydrate oligomer onto specific glycosylation sites on the nascent protein (adapted from http://clemonslab.caltech.edu).

Most of the work concerning N-glycosylation and OST has been done in yeast and only little is known about the processes in plants. In *Arabidopsis*, two homologs of the yeast Stt3, STT3A and STT3B have been identified. *stt3a stt3b* double mutants are gametophytic lethal, but only the *stt3a* single mutant and not *stt3b* alone shows severe global underglycosylation effects (*210*, *211*). This indicates that the two isoforms have partially overlapping, but substrate-specific functions. The *Arabidopsis* homolog of Wbp1, *DEFECTIVE IN GLYCOSYLATION 1* (*DGL1*) has been shown to be actively involved in glycosylation and *dgl1* mutants are embryo-lethal (*212*). *Arabidopsis* has only one homolog of the yeast Ost3p and Ost6p subunits, termed *OST3/6*. OST3/6 is ER-localized and *ost3/6* mutants display severe underglycosylation defects of specific target proteins such as EFR and the cellulose biosynthesis enzyme KORRIGAN1, but not of the receptor-like kinase BRASSINOSTEROID-INSENSITIVE1 (BRI1) (*213*). Furthermore, the

Arabidopsis homologs of Swp1p and Ost2p (*DEFENDER-AGAINST APOPTOTIC DEATH*, *DAD1* and *DAD2*) have been identified but so far no direct involvement in glycosylation has been shown (*214*, *215*).

4 – References

1. M. A. F. Noor, J. L. Feder, Speciation genetics: evolving approaches, *Nature reviews. Genetics* **7**, 851–861 (2006).

2. L. H. Rieseberg, B. K. Blackman, Speciation genes in plants, *Annals of botany* **106**, 439–455 (2010).

3. D. B. Lowry, J. L. Modliszewski, K. M. Wright, C. A. Wu, J. H. Willis, The strength and genetic basis of reproductive isolating barriers in flowering plants, *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3009–3021 (2008).

4. V. D. Vacquier, W. J. Swanson, Selection in the rapid evolution of gamete recognition proteins in marine invertebrates, *Cold Spring Harb Perspect Biol* **3**, a002931–a002931 (2011).

5. T. I. Gossmann *et al.*, Genome wide analyses reveal little evidence for adaptive evolution in many plant species, *Molecular Biology and Evolution* **27**, 1822–1832 (2010).

6. D. C. Presgraves, The molecular evolutionary basis of species formation, *Nature reviews. Genetics*, 1–7 (2010).

7. O. Seehausen *et al.*, Genomics and the origin of species, *Nature reviews. Genetics* **15**, 176–192 (2014).

8. K. Bomblies *et al.*, Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants, *PLoS biology* **5**, e236 (2007).

9. N. L. Clark *et al.*, Coevolution of interacting fertilization proteins, *PLoS Genet* **5**, e1000570 (2009).

10. T. I. Gossmann, M. W. Schmid, U. Grossniklaus, K. J. Schmid, Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*, *Molecular Biology and Evolution* **31**, 574–583 (2014).

11. T. Van der Niet, R. Peakall, S. D. Johnson, Pollinator-driven ecological speciation in plants: new evidence and future perspectives, *Annals of botany* **113**, 199–212 (2014).

12. D. B. Dräger *et al.*, Two genes encoding *Arabidopsis halleri MTP1* metal transport proteins co-segregate with zinc tolerance and account for high *MTP1* transcript levels, *The Plant Journal* **39**, 425–439 (2004).

13. M. Hanikenne *et al.*, Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4, Nature* **453**, 391–395 (2008).

14. M. J. Aranzana *et al.*, Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes, *PLoS Genet* **1** (2005).

15. E. Ogiso-Tanaka *et al.*, Natural Variation of the *RICE FLOWERING LOCUS T 1* Contributes to Flowering Time Divergence in Rice, *PLoS ONE* **8** (2013).

16. E. S. Buckler *et al.*, The Genetic Architecture of Maize Flowering Time, *Science* **325**, 714–718 (2009).

17. L. Fishman, A. L. Sweigart, A. M. Kenney, S. Campbell, Major quantitative trait loci control divergence in critical photoperiod for flowering between selfing and outcrossing species of monkeyflower (*Mimulus*), *New Phytol* **201**, 1498–1507 (2013).

18. K. Bomblies, Doomed lovers: mechanisms of isolation and incompatibility in plants, *Annu Rev Plant Biol* **61**, 109–124 (2010).

19. S. Xu, P. M. Schlüter, U. Grossniklaus, F. P. Schiestl, The Genetic Basis of Pollinator Adaptation in a Sexually Deceptive Orchid, *PLoS Genet* **8**, e1002889 (2012).

20. M. E. Hoballah *et al.*, Single Gene-Mediated Shift in Pollinator Attraction in *Petunia*, *The Plant Cell* **19**, 779–790 (2007).

21. D. L. Des Marais, M. D. Rausher, Parallel Evolution At Multiple Levels In The Origin Of Hummingbird Pollinated Flowers In *Ipomoea*, *Evolution* (2010)

22. K. Schwinn, A Small Family of *MYB*-Regulatory Genes Controls Floral Pigmentation Intensity and Patterning in the Genus *Antirrhinum*, *The Plant Cell* **18**, 831–851 (2006).

23. D. A. Moeller, Geographic structure of pollinator communities, reproductive assurance, and the evolution of self-pollination, *Ecology* **87**, 1510–1522 (2006).

24. C. Tang *et al.*, The evolution of selfing in *Arabidopsis thaliana*, *Science* **317**, 1070–1072 (2007).

25. K. Y. Chen, B. Cong, R. Wing, J. Vrebalov, S. D. Tanksley, Changes in Regulation of a Transcription Factor Lead to Autogamy in Cultivated Tomatoes, *Science* **318**, 643–645 (2007).

26. A. Widmer, C. Lexer, S. Cozzolino, Evolution of reproductive isolation in plants, *Heredity* **102**, 31–38 (2009).

27. L. H. Rieseberg, J. H. Willis, Plant Speciation, *Science* **317**, 910–914 (2007).

28. R. Swanson, A. F. Edlund, D. Preuss, Species Specificity in Pollen-Pistil Interactions, *Annu. Rev. Genet.* **38**, 793–818 (2004).

29. A. Diaz, M. R. Macnair, Pollen tube competition as a mechanism of prezygotic reproductive isolation between *Mimulus nasutus* and its presumed progenitor *M. guttatus*, *New Phytol* **144**, 471–478 (1999).

30. H. F. Smith, Q. D. Clarkson, Cytological studies of interspecific hybridization in *Iris*, subsection *Californiaceae*, *American Journal of Botany* **43**, 582–588 (1956).

31. P. D. Ascher, S. J. Peloquin, Pollen Tube Growth and Incompatibility following Intra- and Inter-Specific Pollinations in *Lilium longiflorum*, *American Journal of Botany* **55**, 1230–1234 (1968).

32. Z. Song, B. Lu, Y. Zhu, J. Chen, Pollen competition between cultivated and wild rice species (*Oryza sativa* and *O. rufipogon*), *New Phytol* **153**, 289–296 (2002).

33. B. R. Montgomery, D. M. Soper, L. F. Delph, Asymmetrical conspecific seed-siring advantage between Silene latifolia and *S. dioica*, *Annals of botany* **105**, 595–605 (2010).

34. J. E. Aagaard, R. D. George, L. Fishman, M. J. Maccoss, W. J. Swanson, Selection on plant male function genes identifies candidates for reproductive isolation of yellow monkeyflowers, *PLoS Genet* **9**, e1003965 (2013).

35. J. Murfett *et al.*, S RNase and Interspecific Pollen Rejection in the Genus *Nicotiana*: Multiple Pollen-Rejection Pathways Contribute to Unilateral Incompatibility between Self-Incompatible and Self-Compatible Species, *The Plant Cell* **8**, 943–958 (1996).

36. S. Okuda *et al.*, Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells, *Nature* **458**, 357–361 (2009).

37. H. Takeuchi, T. Higashiyama, A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*, *PLoS biology* **10**, (2012).

38. R. Swanson, A. F. Edlund, D. Preuss, Species Specificity In Pollen-Pistil Interactions, *Annu. Rev. Genet.* **38**, 793–818 (2004).

39. W. Bateson, Heredity and Variation in Modern Lights, *Darwin and Modern Science*, 85–101 (1909).

40. T. Dobzhansky, Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids, *Genetics* **21**, 113–135 (1936).

41. H. J. Muller, Isolating mechanisms, evolution and temperature, *Biol Symp*, 71–125 (1942).

42. H. A. Orr, The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities, *Genetics* **139**, 1805–1815 (1995).

43. K. Bomblies, D. Weigel, *Arabidopsis*—a model genus for speciation, *Current Opinion in Genetics & Development* **17**, 500–504 (2007).

44. M. Mino, Cell Death Processes during Expression of Hybrid Lethality in Interspecific F1 Hybrid between *Nicotiana gossei* Domin and *Nicotiana tabacum*, *Plant physiology* **130**, 1776–1787 (2002).

45. R. Horn, K. J. Gupta, N. Colombo, Mitochondrion role in molecular basis of cytoplasmic male sterility, *Mitochondrion* (2014).

46. M. R. Hanson, S. Bentolila, Interactions of mitochondrial and nuclear genes that affect male gametophyte development, *The Plant Cell*, 154–69 (2004).

47. G. L. Stebbins, The inviability, weakness, and sterility of interspecific hybrids, *Adv. Genet.* **9**, 147–215 (1958).

48. F. Berger, D. Twell, Germline specification and function in plants, *Annu Rev Plant Biol* **62**, 461–484 (2011).

49. Y. Lou, J. Zhu, Z. Yang, in *Applied Plant Cell Biology*, P. Nick, Z. Opatrny, Eds. (Springer Berlin Heidelberg, 2014), vol. 22, pp. 179–205.

50. D. J. Murphy, The extracellular pollen coat in members of the Brassicaceae: composition, biosynthesis, and functions in pollination, *Protoplasma* **228**, 31–39 (2006).

51. A. H. N. Griffis, N. R. Groves, X. Zhou, I. Meier, Nuclei in motion: movement and positioning of plant nuclei in development, signaling, symbiosis, and disease, *Front Plant Sci* **5**, 129 (2014).

52. R. Yadegari, G. N. Drews, Female gametophyte development, *The Plant Cell* **16**, S133–41 (2004).

53. S. A. Kessler, U. Grossniklaus, She's the boss: signaling in pollen tube reception, *Current opinion in plant biology* **14**, 622–627 (2011).

54. Zinkl GM, Zwiebel BI, Grier DG, P. D, Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine, *Development* **126**, 5431–5440 (1999).

55. T. D. Quilichini, E. Grienenberger, C. J. Douglas, The biosynthesis, composition and assembly of the outer pollen wall: A tough case to crack, *Phytochemistry*, 1–13 (2014).

56. S. J. Hiscock, A. M. Allen, Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus, *New Phytologist* **179**, 286–317 (2008).

57. D. T. Luu, P. Heizmann, C. Dumas, M. Trick, M. Cappadocia, Involvement of *SLR1* genes in pollen adhesion to the stigmatic surface in Brassicaceae, *Sex Plant Reprod*, 1–9 (1997).

58. S. Takayama *et al.*, Isolation and characterization of pollen coat proteins of *Brassica campestris* that interact with *S* locus-related glycoprotein 1 involved in pollen-stigma adhesion, *Proceedings of the National Academy of Sciences* **97**, 3765–3770 (2000).

59. D. T. Luu, D. Marty-Mazyrs, M. Trick, C. Dumas, P. Heizmann, Pollen–Stigma Adhesion in *Brassica* spp Involves SLG and SLR1 Glycoproteins, *The Plant Cell*, 1–13 (1999).

60. S. J. Hiscock, J. Doughty, A. C. Willis, H. G. Dickinson, A 7-kDa pollen coating-borne peptide from *Brassica napus* interacts with *S*-locus glycoprotein and *S*-locus-related glycoprotein, *Planta* **196**, 367–374 (1995).

61. D. T. Luu, E. Passelegue, C. Dumas, P. Heizmann, Pollen-stigma capture is not species discriminant within the Brassicaceae family, *C.R. Acad. Sci. Paris*, 747–755 (1998).

62. M. Hülskamp, S. D. Kopczak, T. F. Horejsi, B. K. Kihl, R. E. Pruitt, Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*, *The Plant Journal* **8**, 703–714 (1995).

63. M. A. Samuel *et al.*, Cellular Pathways Regulating Responses to Compatible and Self-Incompatible Pollen in *Brassica* and *Arabidopsis* Stigmas Intersect at Exo70A1, a Putative Component of the Exocyst Complex, *The Plant Cell* **21**, 2655–2671 (2009).

64. L. A. Chapman, D. R. Goring, Misregulation of phosphoinositides in *Arabidopsis thaliana* decreases pollen hydration and maternal fertility, *Sex Plant Reprod* **24**, 319–326 (2011).

65. J. Mayfield, P. D, Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17, *Nature cell biology*, 1–3 (2000).

66. M. Schein, Rapid Evolution of a Pollen-Specific Oleosin-Like Gene Family from *Arabidopsis thaliana* and Closely Related Species, *Molecular Biology and Evolution* **21**, 659–669 (2004).

67. L. A. Chapman, D. R. Goring, Pollen-pistil interactions regulating successful fertilization in the Brassicaceae, *Journal of experimental botany* **61**, 1987–1999 (2010).

68. Y. Guan, J. Guo, H. Li, Z. Yang, Signaling in pollen tube growth: crosstalk, feedback, and missing links, *Molecular Plant* **6**, 1053–1064 (2013).

69. Y. Qin *et al.*, Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil, *PLoS Genet* **5**, e1000621 (2009).

70. S.-Y. Lin *et al.*, Profiling of translatomes of *in vivo*-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in *Arabidopsis*, *The Plant Cell* **26**, 602–618 (2014).

71. Y. Heslop-Harrison, J. Heslop-Harrison, Germination of Monocolpate Angiosperm Pollen: Evolution of the Actin Cytoskeleton and Wall during Hydration, Activation and Tube Emergence, *Annals of botany*, 385–394 (1991).

72. Y. Wu *et al.*, *Arabidopsis* FIMBRIN5, an actin bundling factor, is required for pollen germination and pollen tube growth, *The Plant Cell* **22**, 3745–3763 (2010).

73. A. J. Castro *et al.*, Electrophoretic profiling and immunocytochemical detection of pectins and arabinogalactan proteins in olive pollen during germination and pollen tube growth, *Annals of botany* **112**, 503–513 (2013).

74. C. J. Elleman, V. E. Franklin-Tong, H. G. Dickinson, Pollination in species with dry stigmas: the nature of the early stigmatic response and the pathway taken by pollen tubes, *New Phytol*, 413–424 (1992).

75. L. Steinhorst, J. Kudla, Calcium - a central regulator of pollen germination and tube growth, *Biochimica et biophysica acta* **1833**, 1573–1581 (2013).

76. L. L. Ge, C. T. Xie, H. Q. Tian, S. D. Russell, Distribution of calcium in the stigma and style of tobacco during pollen germination and tube elongation, *Sex Plant Reprod* **22**, 87–96 (2009).

77. M. Iwano *et al.*, Ca2+ dynamics in a pollen grain and papilla cell during pollination of *Arabidopsis*, *Plant physiology* **136**, 3562–3571 (2004).

78. J. C. Mollet, S. Y. Park, E. A. Nothnagel, E. M. Lord, A lily stylar pectin is necessary for pollen tube adhesion to an in vitro stylar matrix, *The Plant Cell* **12**, 1737–1750 (2000).

79. S. Y. Park *et al.*, A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix, *The Plant Cell* **12**, 151–164 (2000).

80. S. T. Kim, K. Zhang, J. Dong, E. M. Lord, Exogenous free ubiquitin enhances lily pollen tube adhesion to an *in vitro* stylar matrix and may facilitate endocytosis of SCA, *Plant physiology* **142**, 1397–1411 (2006).

81. K. Chae, E. M. Lord, Pollen tube growth and guidance: roles of small, secreted proteins, *Annals of botany* **108**, 627–636 (2011).

82. K. Chae, C. A. Kieslich, D. Morikis, S.-C. Kim, E. M. Lord, A gain-of-function mutation of *Arabidopsis lipid transfer protein 5* disturbs pollen tube tip growth and fertilization, *The Plant Cell* **21**, 3902–3914 (2009).

83. J. Dong, S. T. Kim, E. M. Lord, Plantacyanin plays a role in reproduction in *Arabidopsis*, *Plant physiology* **138**, 778–789 (2005).

84. S. Kim *et al.*, Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism, *Proceedings of the National Academy of Sciences* **100**, 16125–16130 (2003).

85. D. Zhang *et al.*, The pollen receptor kinase *Le*PRK2 mediates growth-promoting signals and positively regulates pollen germination and tube growth, *Plant physiology* **148**, 1368–1379 (2008).

86. D. L. Wengier, M. A. Mazzella, T. M. Salem, S. McCormick, J. P. Muschietti, STIL, a peculiar molecule from styles,

specifically dephosphorylates the pollen receptor kinase *Le*PRK2 and stimulates pollen tube growth *in vitro, BMC Plant Biol.* **10**, 33 (2010).

87. W. Tang, D. Kelley, I. Ezcurra, R. Cotter, S. McCormick, *Le*STIG1, an extracellular binding partner for the pollen receptor kinases *Le*PRK1 and *Le*PRK2, promotes pollen tube growth *in vitro*, *The Plant Journal* **39**, 343–353 (2004).

88. A. Y. Cheung, H. Wang, H.-M. Wu, A Floral Transmitting Tissue-Specific Glycoprotein Attracts Pollen Tubes and Stimulates Their Growth, *Cell* **82**, 282–292 (1995).

89. H.-M. Wu, H. Wang, A. Y. Cheung, A Pollen Tube Growth Stimulatory Glycoprotein Is Deglycosylated by Pollen Tubes and Displays a Glycosylation Gradient in the Flower, *Cell* **82**, 395–403 (1995).

90. C. N. Hancock, L. Kent, B. A. McClure, The stylar 120 kDa glycoprotein is required for *S*-specific pollen rejection in *Nicotiana*, *The Plant Journal* **43**, 716–723 (2005).

91. S. Li *et al.*, *Arabidopsis* COBRA-LIKE 10, a GPI-anchored protein, mediates directional growth of pollen tubes, *The Plant Journal* **74**, 486–497 (2013).

92. T. Dresselhaus, N. Franklin-Tong, Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization, *Molecular Plant* **6**, 1018–1036 (2013).

93. R. Palanivelu, L. Brass, A. F. Edlund, D. Preuss, Pollen tube growth and guidance is regulated by POP2, an *Arabidopsis* gene that controls GABA levels, *Cell* **114**, 47–59 (2003).

94. G.-H. Yu, M.-X. Sun, Deciphering the Possible Mechanism of GABA in Tobacco Pollen Tube Growth and Guidance, *psb* **2**, 393–395 (2007).

95. E. Michard *et al.*, Glutamate receptor-like genes form Ca2+ channels in pollen tubes and are regulated by pistil D-serine, *Science* **332**, 434–437 (2011).

96. A. M. Prado, R. Colaço, N. Moreno, A. C. Silva, J. A. Feijó, Targeting of pollen tubes to ovules is dependent on nitric oxide (NO) signaling, *Molecular Plant* **1**, 703–714 (2008).

97. S. C. Baker, K. Robinson-Beers, J. M. Villanueva, J. C. Gaiser, C. S. Gasser, Interactions among genes regulating ovule development in *Arabidopsis thaliana*, *Genetics* **145**, 1109–1124 (1997).

98. Y. Lu *et al.*, Pollen tubes lacking a pair of K+ transporters fail to target ovules in *Arabidopsis*, *The Plant Cell* **23**, 81–93 (2011).

99. Y. Guan, J. Lu, J. Xu, B. McClure, S. Zhang, Two Mitogen-Activated Protein Kinases, MPK3 and MPK6, Are Required for Funicular Guidance of Pollen Tubes in *Arabidopsis*, *Plant physiology* **165**, 528–533 (2014).

100. M. K. Kandasamy, J. B. Nasrallah, M. E. Nasrallah, Pollen-pistil interactions and developmental regulation of pollen tube growth in *Arabidopsis*, *Development* **120**, 3405–3418 (1994).

101. T. Higashiyama *et al.*, Pollen tube attraction by the synergid cell, *Science* **293**, 1480–1483 (2001).

102. Y. Ling *et al.*, Mutation in SUMO E3 ligase, *SIZ1*, disrupts the mature female gametophyte in *Arabidopsis*, *PLoS ONE* **7** (2012).

103. Y. H. Chen *et al.*, The central cell plays a critical role in pollen tube guidance in *Arabidopsis*, *The Plant Cell* **19**, 3563–3577 (2007).

104. K. K. Shimizu, K. Okada, Attractive and repulsive interactions between female and male gametophytes in *Arabidopsis* pollen tube guidance, *Development* **127**, 4511–4518 (2000).

105. H. Wang, L. C. Boavida, M. Ron, S. McCormick, Truncation of a protein disulfide isomerase, *PDIL2-1*, delays embryo sac maturation and disrupts pollen tube guidance in *Arabidopsis thaliana*, *The Plant Cell* **20**, 3300–3311 (2008).

106. R. D. Kasahara, M. F. Portereiko, L. Sandaklie-Nikolova, D. S. Rabiger, G. N. Drews, *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*, *The Plant Cell* **17**, 2981–2992 (2005).

107. M. Alandete-Saez, M. Ron, S. McCormick, *GEX3*, expressed in the male gametophyte and in the egg cell of *Arabidopsis thaliana*, is essential for micropylar pollen tube guidance and plays a role during early embryogenesis, *Molecular Plant* **1**, 586–598 (2008).

108. M. L. Marton, S. Cordts, J. Broadhvest, T. Dresselhaus, Micropylar pollen tube guidance by *egg apparatus 1* of maize, *Science* **307**, 573–576 (2005).

109. T. Higashiyama *et al.*, Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*, *Plant physiology* **142**, 481–491 (2006).

110. S. Okuda *et al.*, Acquisition of LURE-binding activity at the pollen tube tip of *Torenia fournieri, Molecular Plant* **6**, 1074–1090 (2013).

111. M. W. Jones-Rhoades, J. O. Borevitz, D. Preuss, Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins, *PLoS Genet* **3**, 1848–1861 (2007).

112. M. L. Marton, A. Fastner, S. Uebler, T. Dresselhaus, Overcoming Hybridization Barriers by the Secretion of the Maize Pollen Tube Attractant *Zm*EA1 from *Arabidopsis* Ovules, *Current biology* (2012).

113. S. Uebler, T. Dresselhaus, M.-L. Márton, Species-specific interaction of EA1 with the maize pollen tube apex, *Plant Signal Behav* **8** (2013).

114. J. Liu *et al.*, Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male-female attraction in *Arabidopsis*, *Current biology* **23**, 993–998 (2013).

115. H. J. Li *et al.*, *POD1* regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis*, *The Plant Cell* **23**, 3288–3302 (2011).

116. Y. Leshem, C. Johnson, V. Sundaresan, Pollen tube entry into the synergid cell of *Arabidopsis* is observed at a site distinct from the filiform apparatus, *Plant Reprod* **26**, 93–99 (2013).

117. Q. A. Ngo, H. Vogler, D. S. Lituiev, A. Nestorova, U. Grossniklaus, A Calcium Dialog Mediated by the *FERONIA* Signal Transduction Pathway Controls Plant Sperm Delivery, *Dev. Cell* **29**, 491–500 (2014).

118. N. Huck, J. M. Moore, M. Federer, U. Grossniklaus, The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception, *Development* **130**, 2149–2159 (2003).

119. N. Rotman *et al.*, Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*, *Current biology* **13**, 432–436 (2003).

120. H. Lindner, L. M. Muller, A. Boisson-Dernier, U. Grossniklaus, *Cr*RLK1L receptor-like kinases: not just another brick in the wall, *Current opinion in plant biology* **15**, 659–669 (2012).

121. J. M. Escobar-Restrepo *et al.*, The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception, *Science* **317**, 656–660 (2007).

122. A. Boisson-Dernier, S. A. Kessler, U. Grossniklaus, The walls have ears: the role of plant *Cr*RLK1Ls in sensing and transducing extracellular signals, *Journal of experimental botany* **62**, 1581–1591 (2011).

123. V. R. J. L. W. E. G. Kaul, Early events in the embryo sac after intraspecific and interspecific pollinations in *Rhododendron kawakamii* and *R. retusum*, *Canadian Journal of Botany* **64**, 282–291 (1986).

124. E. G. Williams, V. Kaul, J. L. Rouse, B. F. Palser, Overgrowth of Pollen Tubes in Embryo Sacs of *Rhododendron* Following Interspecific Pollinations, *Aust. J. Bot.* **34**, 413–423 (1986).

125. M. Haruta, G. Sabat, K. Stecker, B. B. Minkoff, M. R. Sussman, A peptide hormone and its receptor protein kinase regulate plant cell expansion, *Science* **343**, 408–411 (2014).

126. A. Capron *et al.*, Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene, *The Plant Cell* **20**, 3038–3049 (2008).

127. S. A. Kessler *et al.*, Conserved molecular components for pollen tube reception and fungal invasion, *Science* **330**, 968–971 (2010).

128. H. Lindner *et al.*, SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing, *Genetics* (2012).

129. N. Rotman, M. Gourgues, A. E. Guitton, J. E. Faure, F. Berger, A dialogue between the *SIRENE* pathway in synergids and the fertilization independent seed pathway in the central cell controls male gamete release during double fertilization in *Arabidopsis*, *Molecular Plant* **1**, 659–666 (2008).

130. A. Boisson-Dernier, S. Frietsch, T. H. Kim, M. B. Dizon, J. I. Schroeder, The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition, *Current biology* **18**, 63–68 (2008).

131. Q. Duan, D. Kita, C. Li, A. Y. Cheung, H. M. Wu, FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development, *Proceedings of the National Academy of Sciences* **107**, 17821–17826 (2010).

132. Q. Duan *et al.*, Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in *Arabidopsis*, *Nat Commun* **5**, 3129 (2014).

133. S. Amien *et al.*, Defensin-like *Zm*ES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1, *PLoS biology* **8** (2010).

134. S. Wolf, K. Hematy, H. Hofte, Growth control and cell wall signaling in plants, *Annu Rev Plant Biol* **63**, 381–407 (2012).

135. M. Woriedh *et al.*, External application of gametophyte-specific *Zm*PMEI1 induces pollen tube burst in maize, *Plant Reprod* **26**, 255–266 (2013).

136. A. Boisson-Dernier *et al.*, *ANXUR* receptor-like kinases coordinate cell wall integrity with growth at the pollen tube tip via NADPH oxidases, *PLoS biology* **11** (2013).

137. A. Boisson-Dernier *et al.*, Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge, *Development* **136**, 3279–3288 (2009).

138. S. Miyazaki *et al.*, *ANXUR1* and *2*, sister genes to *FERONIA*/*SIRENE*, are male factors for coordinated fertilization, *Current biology* **19**, 1327–1331 (2009).

139. M. M. Kanaoka, K. U. Torii, *FERONIA* as an upstream receptor kinase for polar cell growth in plants, *Proceedings of the National Academy of Sciences* **107**, 17461–17462 (2010).

140. M. Schiøtt *et al.*, A plant plasma membrane Ca2+ pump is required for normal pollen tube growth and fertilization, *Proceedings of the National Academy of Sciences* **101**, 9502–9507 (2004).

141. A. R. Leydon *et al.*, Three MYB transcription factors control pollen tube differentiation required for sperm release, *Current biology* **23**, 1209–1214 (2013)

142. Y. Liang *et al.*, MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization, *PLoS Genet* **9** (2013).

143. X. Zhou, I. Meier, Efficient plant male fertility depends on vegetative nuclear movement mediated by two families of plant outer nuclear membrane proteins, *Proceedings of the National Academy of Sciences* (2014).

144. T. Igawa, Y. Yanagawa, S.-Y. Miyagishima, T. Mori, Analysis of gamete membrane dynamics during double fertilization of *Arabidopsis*, *J. Plant Res.* **126**, 387–394 (2013).

145. T. Mori, T. Igawa, G. Tamiya, S.-Y. Miyagishima, F. Berger, Gamete attachment requires GEX2 for successful fertilization in *Arabidopsis*, *Current biology* **24**, 170–175 (2014).

146. N. Inoue, M. Ikawa, A. Isotani, M. Okabe, The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs, *Nature* **434**, 234–238 (2005).

147. T. Mori, H. Kuroiwa, T. Higashiyama, T. Kuroiwa, *GENERATIVE CELL SPECIFIC 1* is essential for angiosperm fertilization, *Nature cell biology* **8**, 64–71 (2006).

148. K. von Besser, A. C. Frank, M. A. Johnson, D. Preuss, *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization, *Development* **133**, 4761–4769 (2006).

149. Y. Liu *et al.*, The conserved plant sterility gene *HAP2* functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes, *Genes & Development* **22**, 1051–1068 (2008).

150. S. Sprunck *et al.*, Egg cell-secreted EC1 triggers sperm cell activation during double fertilization, *Science* **338**, 1093–1097 (2012).

151. Y. Leshem *et al.*, Molecular characterization of the *glauce* mutant: a central cell-specific function is required for double fertilization in *Arabidopsis*, *The Plant Cell* **24**, 3264–3277 (2012).

152. F. Yu *et al.*, ANK6, a mitochondrial ankyrin repeat protein, is required for male-female gamete recognition in *Arabidopsis thaliana*, *Proceedings of the National Academy of Sciences* **107**, 22332–22337 (2010).

153. M. Spielman, R. J. Scott, Polyspermy barriers in plants: from preventing to promoting fertilization, *Sex Plant Reprod* **21**, 53–65 (2008).

154. K. M. Beale, A. R. Leydon, M. A. Johnson, Gamete Fusion Is Required to Block Multiple Pollen Tubes from Entering an *Arabidopsis* Ovule, *Current biology* (2012).

155. R. D. Kasahara *et al.*, Fertilization Recovery after Defective Sperm Cell Release in *Arabidopsis*, *Current biology* (2012).

156. D. Maruyama *et al.*, Independent control by each female gamete prevents the attraction of multiple pollen tubes, *Dev. Cell* **25**, 317–323 (2013).

157. R. Völz, J. Heydlauff, D. Ripper, L. von Lyncker, R. Groß-Hardt, Ethylene signaling is required for synergid degeneration and the establishment of a pollen tube block, *Dev. Cell* **25**, 310–316 (2013).

158. M. Iwano, S. Takayama, Self/non-self discrimination in angiosperm self-incompatibility, *Current opinion in plant biology* **15**, 78–83 (2012).

159. J. C. Stein, R. Dixit, M. E. Nasrallah, J. B. Nasrallah, *SRK*, the stigma-specific *S* locus receptor kinase of Brassica, is targeted to the plasma membrane in transgenic tobacco, *The Plant Cell* **8**, 429–445 (1996).

160. C. R. Schopfer, M. E. Nasrallah, J. B. Nasrallah, The male determinant of self-incompatibility in *Brassica*, *Science* **286**, 1697–1700 (1999).

161. T. Takasaki *et al.*, The *S* receptor kinase determines self-incompatibility in *Brassica* stigma, *Nature* **403**, 913–916 (2000).

162. S. Takayama *et al.*, Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility, *Nature* **413**, 534–538 (2001).

163. M. Kusaba *et al.*, Self-incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana, The Plant Cell* **13**, 627–643 (2001).

164. M. Iwano *et al.*, Immunohistochemical studies on translocation of pollen *S*-haplotype determinant in self-incompatibility of *Brassica rapa*, *Plant and Cell Physiology* **44**, 428–436 (2003).

165. R. Ivanov, I. Fobis-Loisy, T. Gaude, When no means no: guide to Brassicaceae self-incompatibility, *Trends Plant Sci.* **15**, 387–394 (2010).

166. T. Tsuchimatsu *et al.*, Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene, *Nature* **464**, 1342–1346 (2010).

167. B. McClure, F. Cruz-Garcia, C. Romero, Compatibility and incompatibility in *S*-RNase-based systems, *Annals of botany* **108**, 647–658 (2011).

168. B. McClure, B. Mou, S. Canevascini, R. Bernatzky, A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*, *Proceedings of the National Academy of Sciences* **96**, 13548–13553 (1999).

169. F. Cruz-Garcia, C. Nathan Hancock, D. Kim, B. McClure, Stylar glycoproteins bind to *S*-RNase in vitro, *The Plant Journal* **42**, 295–304 (2005).

170. D. T. Luu, X. Qin, D. Morse, M. Cappadocia, *S*-RNase uptake by compatible pollen tubes in gametophytic self-

incompatibility, *Nature* **407**, 649–651 (2000).

171. M. J. Wheeler *et al.*, Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*, *Nature* **459**, 992–995 (2009).

172. S. G. Thomas, V. E. Franklin-Tong, Self-incompatibility triggers programmed cell death in *Papaver* pollen, *Nature* **429**, 305–309 (2004).

173. M. Bosch, V. E. Franklin-Tong, Self-incompatibility in *Papaver*: signalling to trigger PCD in incompatible pollen, *Journal of experimental botany* **59**, 481–490 (2008).

174. D. Lewis, L. K. Crowe, Unilateral interspecific incompatibility in flowering plants, *Heredity* **12**, 233–256 (1958).

175. R. T. Chetelat, J. W. Deverna, Expression of unilateral incompatibility in pollen of *Lycopersicon pennellii* is determined by major loci on chromosomes 1, 6 and 10, *Theor. Appl. Genet.* **82**, 704–712 (1991).

176. P. A. Covey *et al.*, Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade, *The Plant Journal* **64**, 367–378 (2010).

177. W. Li, R. T. Chetelat, A pollen factor linking inter- and intraspecific pollen rejection in tomato, *Science* **330**, 1827–1830 (2010).

178. Y. Takada *et al.*, Involvement of MLPK Pathway in Intraspecies Unilateral Incompatibility Regulated by a Single Locus with Stigma and Pollen Factors, *G3 (Bethesda)* **3**, 719–726 (2013).

179. A. Tovar-Méndez *et al.*, Restoring pistil-side self-incompatibility factors recapitulates an interspecific reproductive barrier between tomato species, *The Plant Journal* **77**, 727–736 (2014).

180. H. J. Li, W. C. Yang, Emerging role of ER quality control in plant cell signal perception, *Protein Cell* **3**, 10–16 (2012).

181. R. Knauer, L. Lehle, The oligosaccharyltransferase complex from yeast, *Biochimica et biophysica acta* **1426**, 259–273 (1999).

182. M. G. Buffone *et al.*, Heads or tails? Structural events and molecular mechanisms that promote mammalian sperm acrosomal exocytosis and motility, *Mol. Reprod. Dev.* **79**, 4–18 (2012).

183. P. M. Wassarman, E. S. Litscher, Mammalian fertilization: the egg's multifunctional zona pellucida, *Int. J. Dev. Biol.* **52**, 665–676 (2008).

184. G. F. Clark, Molecular models for mouse sperm-oocyte binding, *Glycobiology* **21**, 3–5 (2011).

185. W. J. Swanson, V. D. Vacquier, The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule, *Proceedings of the National Academy of Sciences* **94**, 6724–6729 (1997).

186. R. Buschges *et al.*, The barley *Mlo* gene: a novel control element of plant pathogen resistance, *Cell* **88**, 695–705 (1997).

187. R. A. Bhat, M. Miklis, E. Schmelzer, P. Schulze-Lefert, R. Panstruga, Recruitment and interaction dynamics of plant penetration resistance components in a plasma membrane microdomain, *Proceedings of the National Academy of Sciences* **102**, 3135–3140 (2005).

188. R. J. O'Connell, R. Panstruga, Tete a tete inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes, *New Phytol* **171**, 699–718 (2006).

189. F. G. Malinovsky, J. U. Fangel, W. G. T. Willats, The role of the cell wall in plant immunity, *Front Plant Sci* **5**, 178 (2014).

190. L. Gómez-Gómez, T. Boller, FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*, *Molecular cell* **5**, 1003–1011 (2000).

191. C. Zipfel *et al.*, Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation, *Cell* **125**, 749–760 (2006).

192. G. Buist, A. Steen, J. Kok, O. P. Kuipers, LysM, a widely distributed protein motif for binding to (peptido)glycans, *Mol. Microbiol.* **68**, 838–847 (2008).

193. T. Ohnuma, S. Onaga, K. Murata, T. Taira, E. Katoh, LysM domains from *Pteris ryukyuensis* chitinase-A: a stability study and characterization of the chitin-binding site, *The Journal of biological chemistry* **283**, 5178–5187 (2008).

194. T. Liu *et al.*, Chitin-induced dimerization activates a plant immune receptor, *Science* **336**, 1160–1164 (2012).

195. A. A. Gust, R. Willmann, Y. Desaki, H. M. Grabherr, T. Nürnberger, Plant LysM proteins: modules mediating symbiosis and immunity, *Trends Plant Sci.* **17**, 495–502 (2012).

196. T. Shimizu *et al.*, Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice, *The Plant Journal* **64**, 204–214 (2010).

197. A. Miya *et al.*, CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*, *Proceedings of the National Academy of Sciences* **104**, 19613–19618 (2007).

198. R. Willmann *et al.*, *Arabidopsis* lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection, *Proceedings of the National Academy of Sciences* **108**, 19824–19829 (2011).

199. S. Radutoiu *et al.*, Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases, *Nature* **425**, 585–592 (2003).

200. L. Mulder, B. Lefebvre, J. Cullimore, A. Imberty, LysM domains of *Medicago truncatula* NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors, *Glycobiology* **16**, 801–809 (2006).

201. E. Limpens *et al.*, LysM domain receptor kinases regulating rhizobial Nod factor-induced infection, *Science* **302**, 630–633 (2003).

202. V. Zhukov *et al.*, The pea *Sym37* receptor kinase gene controls infection-thread initiation and nodule development, *Mol. Plant Microbe Interact.* **21**, 1600–1608 (2008).

203. R. Op den Camp *et al.*, LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*, *Science* **331**, 909–912 (2011).

204. S. Radutoiu *et al.*, LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range, *EMBO J.* **26**, 3923–3935 (2007).

205. M. Schwarz, R. Knauer, L. Lehle, Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit, *FEBS letters* **579**, 6564–6568 (2005).

206. B. L. Schulz *et al.*, Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency, *Proceedings of the National Academy of Sciences* **106**, 11061–11066 (2009).

207. W. Breuer, R. A. Klein, B. Hardt, A. Bartoschek, E. Bause, Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser, *FEBS letters* **501**, 106–110 (2001).

208. J. Breitling, M. Aebi, N-linked protein glycosylation in the endoplasmic reticulum, *Cold Spring Harb Perspect Biol* **5** (2013).

209. M. Aebi, N-linked protein glycosylation in the ER, *Biochimica et biophysica acta* **1833**, 2430–2437 (2013).

210. H. Koiwa *et al.*, The STT3a subunit isoform of the *Arabidopsis* oligosaccharyltransferase controls adaptive responses to salt/osmotic stress, *The Plant Cell* **15**, 2273–2284 (2003).

211. H. Haweker *et al.*, Pattern recognition receptors require N-glycosylation to mediate plant immunity, *The Journal of biological chemistry* **285**, 4629–4636 (2010).

212. O. Lerouxel *et al.*, Mutants in *DEFECTIVE GLYCOSYLATION*, an *Arabidopsis* homolog of an oligosaccharyltransferase complex subunit, show protein underglycosylation and defects in cell differentiation and growth, *The Plant Journal* **42**, 455–468 (2005).

213. A. Farid *et al.*, Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses, *Plant physiology* **162**, 24–38 (2013).

214. A. Danon, V. I. Rotari, A. Gordon, N. Mailhac, P. Gallois, Ultraviolet-C overexposure induces programmed cell death in *Arabidopsis*, which is mediated by caspase-like activities and which can be suppressed by caspase inhibitors, p35 and Defender against Apoptotic Death, *The Journal of biological chemistry* **279**, 779–787 (2004).

215. M. A. Johnson *et al.*, *Arabidopsis* hapless mutations define essential gametophytic functions, *Genetics* **168**, 971–982 (2004).

5 – CrRLK1L receptor-like kinases: not just another brick in the wall

Heike Lindner[1], Lena Maria Müller[1], Aurélien Boisson-Dernier[*] and Ueli Grossniklaus[*]

Institute of Plant Biology & Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

[1]These authors contributed equally to this work.

[*]Authors for correspondence (e-mail: aboisson@access.uzh.ch, grossnik@botinst.uzh.ch)

This review has been published in the present form in *Current opinion in Plant Biology*, 15, 659-669 (2012)

Abstract

In plants, receptor-like kinases regulate many processes during reproductive and vegetative development. The *Arabidopsis* subfamily of *Catharanthus roseus* RLK1-like kinases (CrRLK1Ls) comprises 17 members with a putative extracellular carbohydrate-binding malectin-like domain. Only little is known about the functions of these proteins, although mutant analyses revealed a role during cell elongation, polarized growth, and fertilization. However, the molecular nature of the underlying signal transduction cascades remains largely unknown. CrRLK1L proteins are also involved in biotic and abiotic stress response. It is likely that carbohydrate-rich ligands transmit a signal, which could originate from cell wall components, an arriving pollen tube, or a pathogen attack. Thus, post-translational modifications could be crucial for CrRLK1L signal transduction and ligand binding.

Introduction

Plant cells are directing the synthesis and deposition of an extracellular matrix known as the cell wall. This rigid yet dynamic cell wall fulfills various functions that are central to plant growth and development, such as withholding positive and negative pressure, providing cell-to-cell adhesion, directing growth, or constituting the first contact barrier

between the cell and its neighboring cells, as well as abiotic stresses and invading microorganisms. Despite their diversity between plant species, organs, cell types, or even microdomains (*1*), all primary cell walls are made of a complex mixture of carbohydrate components. These typically include cellulose, the main load-bearing structure, which is cross-linked with hemicelluloses and embedded in a matrix of pectins and secreted (glyco)proteins. The cells synthesize primary cell walls during growth, which depends on the balance between loosening/deformation of the pre-existing cell wall and addition of new cell wall material at the site of growth (*2*). This coordination requires the growing cell to be informed of any perturbations that impact the properties of its cell wall. The latter is therefore able to modulate the cell's activities, such that the cell can adapt during growth. How a plant cell controls the assembly and remodeling of its cell wall during growth while maintaining its mechanical integrity is one of the most fascinating questions in plant biology (*3*). In yeast and fungi, two other organisms with cell walls, a conserved cell wall integrity (CWI) maintenance system has been characterized in great detail (*4,5*). In plants, the existence of cell wall sensing mechanisms was revealed by studies showing that mutations affecting cell wall synthesis resulted in the activation of hormone signaling pathways and increased resistance to pathogens (e.g. (*6,7*)). The plant receptor-like kinase (RLK) superfamily has been shown to mediate the sensing to many extracellular cues to regulate intracellular activities (*8,9*). RLKs are typically composed of an intracellular serine/threonine kinase domain, a transmembrane domain, and a varied extracellular domain (ECD). The functions of a subset of these RLKs, namely the wall-associated kinases (WAKs), Pro-rich extensin-like receptor kinases (PERKs), lectin receptor kinase LecRKs and some leucine-rich repeat (LRR) RLKs, appear to be tightly linked to cell wall synthesis, remodeling, and sensing (reviewed in (*3,10-12*)). Regarding cell wall surveillance mechanisms, the *Catharantus roseus* RLK1-like (*CrRLK1L*) subfamily with its 17 members (*At5g38990*, *At5g39000*, *At5g39020*, *At5g39030*, *FER – At3g51550*, *ANX2 –At5g28680*, *ANX1 – At3g04690*, *HERK2 – At1g30570*, *At4g39110*, *At2g21480*, *At5g61350*, *THE1 – At5g54380*, *At2g39360*, *At5g59700*, *HERK1 – At3g46290*, *At5g24010*, *At2g23200*) has received increasing attention over the past 6 years (*13-15*). CrRLK1 proteins have an ECD with two domains showing limited homology to the carbohydrate-binding domain of the animal Malectin protein (*16*). Six out of 17 subfamily members have been proposed to function in coordinating cell growth, cell-cell communication, and cell wall remodeling during both vegetative and reproductive development. Here, we review recent findings

concerning down-stream targets of the CrRLK1Ls and discuss their putative role in responses to abiotic and biotic stresses and the putative importance of post-translational modifications for these RLKs.

THE ROLE OF CrRLK1Ls IN CELL ELONGATION

FERONIA, *HERCULES1* and *THESEUS1* control cell elongation during vegetative development

Three members of the CrRLK1L subfamily, encoded by *THESEUS1* (*THE1*) and *HERCULES1* (*HERK1*), both named after Greek mythological figures, and *FERONIA* (*FER*), named after an Etruscan goddess of fertility, are plasma membrane-localized receptor-like Ser/Thr-kinases (*17-19*) with some common but also distinct functions during plant development.

THE1 was discovered as suppressor of the cellulose synthase-deficient mutant *procuste1* (*cesA6*$^{prc1-1}$), partially rescuing its short hypocotyl phenotype when grown in the dark (Figure 1a) (*18*). This suppression is not due to a recovery of cellulose biosynthesis in the *prc1/the1* double mutant. Rather, sensing of perturbations in cell wall composition seems affected since the *the1* mutation has no apparent phenotype in a cell-wall intact context. This was the first evidence that a defect in cellulose synthesis did not restrict cell growth physically, but that cell elongation is actively inhibited by *THE1*-dependent sensing of cell wall perturbations. Moreover, *THE1* is required for lignin accumulation that is either caused by mutations in the cellulose-synthase complex (*18*) or by the cellulose synthesis inhibitor isoxaben (*20*). Isoxaben treatment also induces the accumulation of reactive oxygen species (ROS) in wild-type seedlings, but this happens neither in *the1* nor in *rbohD/rbohF* mutants that affect the ROS-producing NADPH oxidase *(20*). These studies indicate that if CWI is not sustained, THE1 activates a signal transduction pathway leading to ROS production, growth inhibition, and lignin accumulation.

In contrast to its inhibitory effect on cellular growth in cell wall damaged contexts, *THE1* was found to be required for cell elongation during vegetative growth together with *HERK1* and *FER* (Figure 1a) (*19*). All three genes were shown to be up-regulated after treatment with brassinosteroid (BR) (*21*), a plant hormone involved in many growth and developmental processes (*22*). Additionally, all three genes were shown to be

plasma membrane-localized and are widely expressed in most vegetative tissues such as leaves, stems and roots (Figure 1b) (*17-19*), with a stronger expression in regions with elongating cells (*19*). *THE1* and *HERK1* seem to act redundantly since the single null mutants have no obvious vegetative phenotype. However, the *the1/herk1* double mutant displays cell elongation defects in leaves and leaf petioles with an even stronger phenotype in the *the1/herk1/herk2* triple mutant (*23*). Homozygous *fer* single mutants show a comparable cell elongation defect, leading to the hypothesis that FER could act as a co-receptor of THE1/HERK RLKs. In addition, similar genes are mis-expressed in the respective mutants, further supporting that *THE1*/*HERK* and *FER* may act in a common pathway (*19*).

Despite the transcriptional increase of the three members of the *CrRLK1L* subfamily after BR treatment, the *THE1*/*HERK* pathway regulating cell elongation seems to be largely independent of BRs because only a subset of BR influenced genes are affected in the *the1/herk1* mutant (*19*). In contrast, it was shown that *fer* null mutants display an imbalance of endogenous ethylene and BRs (*24*). Whereas hypocotyls of light-grown *fer* mutants show an enhanced BR response, dark-grown etiolated *fer* seedlings are partially BR-insensitive and displayed an enhanced ethylene response, leading to shorter hypocotyls in the mutant (Figure 1a).

FERONIA, *ANXUR1* and *ANXUR2* maintain cell wall integrity to sustain polar growth

In addition to its role during cell elongation in leaves and hypocotyls (*19*), Duan and colleagues (*25*) revealed an interaction between FER and ROPGEF1 in yeast and root protoplast cells. ROPGEFs belong to the guanine exchange factor family (*26,27*), activating Rho-like GTPases, monomeric GTP-binding proteins that are known as RAC/ROPs in plants. These RAC/ROPs accomplish various signaling functions that regulate development and polar growth (*28*).

Interestingly, the disruption of *FER* disturbs the polar growth of root hairs, as *fer* mutant roots carry arrested, collapsed, short, and bursting root hairs (Figure 1a) (*25*). This phenotype is similar to that of the *root hair defective2 (rhd2)* mutant, in which the ROS-producing NADPH oxidase *RbohC* is affected (*29-31*). Reduced levels of activated RAC/ROP in *fer* null mutants support the role of FER as an upstream regulator of ROPs (*32*). Furthermore, reduced levels of ROS in *fer* compared to wild-type roots indicates

that the interaction of FER and ROPGEFs, with the subsequent activation of ROPs, result in NADPH oxidase-dependent ROS production.

Intriguingly, *FER*'s two closest homologs in *Arabidopsis* are preferentially expressed in pollen tubes (Figure 1b) – the male gametophytes that, like root hairs, elongate polarly – and are localized to the plasma membrane of the pollen tube tip (*33,34*). Thus, they were named after the male consort of Feronia, *ANXUR1* (*ANX1*) and *ANXUR2* (*ANX2*). They share 85.6% amino acid level identity and work redundantly since only double mutant plants showed a severe male sterility phenotype. Sterility occurs because *anx1/anx2* double mutant pollen tubes lose their cellular integrity and burst during growth between the stigma and the style, preventing them to reach and fertilize the female gametes (Figure 1a) (*33,34*). Because of the similarity between the *fer* root hair and *anx1/anx2* pollen tube phenotypes, it would be important to investigate whether ANX1/2 regulate NADPH oxidase-dependent ROS production in pollen tubes as FER does in root hairs. Nonetheless, FER and ANX1/2 appear to function at the tip of polarly growing cells to prevent loss of CWI during rapid growth.

CrRLKL1s IN CELL-CELL COMMUNICATION

FERONIA triggers pollen tube discharge during fertilization

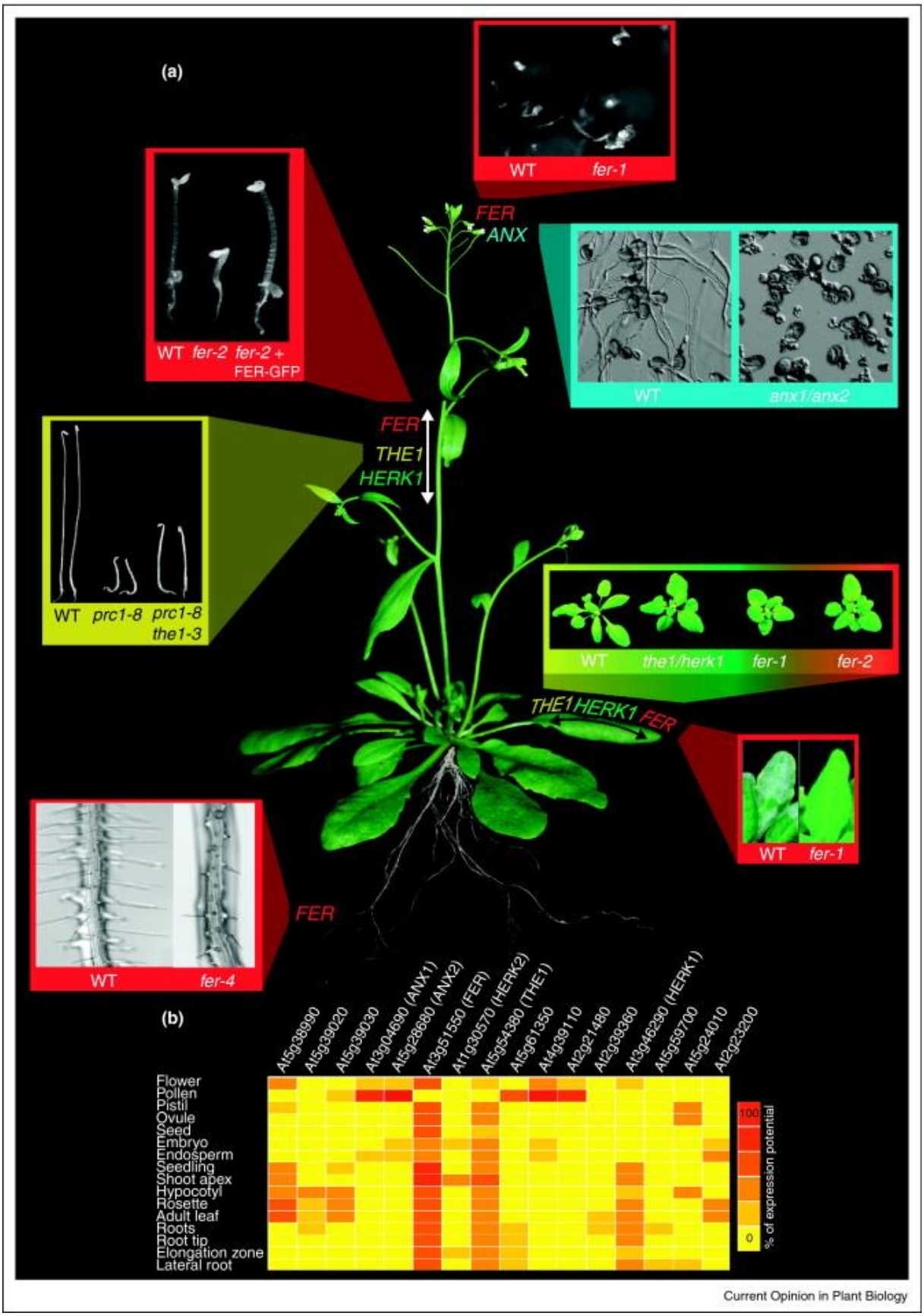
Communication between cells is a crucial process in plant development, reproduction, and defense. The communication between male and female tissues is important during every step of pollination and fertilization: communication between stigma cells and pollen grains leads to pollen hydration and germination of the pollen tubes (*35*). Importantly, after growing through the style and transmitting tract pollen tubes are guided towards the ovules by small cysteine-rich proteins that are secreted by the synergid cells (*36,37*). The two female gametes, the egg and central cell, are also involved in this process but their role is less well understood (*38,39*). Importantly, cell-cell communication is required during the final stages of reproduction, when the pollen tube enters the degenerating synergid, ruptures and releases the two sperm cells to effect double fertilization (*17,40*).

The allelic female gametophytic *fer* and *sirene* mutants – hereafter referred to as *fer* – were identified because they disrupt the communication process between male and female cells during pollen tube reception (*41,42*). *FER* is – with the exception of mature

pollen – expressed throughout the whole plant (Figure 1b) and strongly accumulates in the female gametophyte at the filiform apparatus, a specialized cell wall structure at the basal end of the synergid cells (*17,43*). In *fer* mutant ovules, the pollen tube does not rupture after entering the female gametophyte, but instead continues to grow in the embryo sac, leading to a pollen tube overgrowth phenotype (Figure 1a) (*41,42*). Consequently, the *fer* mutant embryo sacs remain unfertilized. In addition, *fer* mutant ovules can attract supernumerary pollen tubes. The mutation does not affect the development or identity of female gametophytic cells, thus only the communication process between male and female gametophytes appears impaired. Hypothetically, FER could recognize a ligand secreted or carried by the pollen tube. Alternatively, FER could interact with a ligand originating from female tissues, which could reflect a change or modification of the cell wall structure caused by the arriving pollen tube.

Whereas *ANX1/2* seem to be required to sustain pollen tube growth and prevent bursting within female tissues, *FER* inhibits pollen tube growth and positively regulates pollen tube discharge after pollen tube arrival in the synergid. Therefore, it is possible that upon pollen tube arrival the *ANX1/2*-dependent pathway gets inactivated in pollen, while the *FER* signaling cascade becomes activated in the female gametophyte to prepare pollen tube discharge and fertilization. It has been proposed that during pollen tube reception FER and ANX1/2 could compete for the same ligand, a competition that would result in inactivation of the ANX1/2-dependent pollen tube integrity and the activation of the FER-mediated pollen tube recognition (*44*).

After the discovery of *FER*, several other mutants displaying a similar pollen tube overgrowth phenotype in the embryo sac have been identified: *nortia* (*nta*), which will be discussed below (*45*); *lorelei* (*lre*), a mutation in a glycosylphosphatidylinositol (GPI)-anchored protein expressed in the embryo sac (*46*); *scylla* (*syl*), which displays, in addition to pollen tube overgrowth, fertilization-independent endosperm development (*47*), and the mutant *abstinence by mutual consent* (*amc*) affecting a gene encoding a peroxin, where the failure in pollen tube growth arrest is only observed when both male and female gametophytes carry the mutation (*48*). The latest mutant with disrupted pollen tube reception and *fer*-like pollen tube overgrowth is *turan* (*tun*), named after an Etruscan goddess of love and fertility, affecting a gene encoding a UDP-Glycosyltransferase superfamily protein (*49*). The exact role these genes play in the *FER* signal transduction pathway remains to be elucidated.


(a)
WT fer-1
FER
ANX
WT fer-2 fer-2 + FER-GFP
WT anx1/anx2
FER
THE1
HERK1
WT prc1-8 prc1-8 the1-3
WT the1/herk1 fer-1 fer-2
THE1 HERK1 FER
WT fer-1
FER
WT fer-4
(b)
At5g38990
At5g39020
At5g39030
At3g04690 (ANX1)
At5g28680 (ANX2)
At3g51550 (FER)
At1g30570 (HERK2)
At5g54380 (THE1)
At5g61350
At4g39110
At2g21480
At2g39360
At3g46290 (HERK1)
At5g59700
At5g24010
At2g23200
Flower
Pollen
Pistil
Ovule
Seed
Embryo
Endosperm
Seedling
Shoot apex
Hypocotyl
Rosette
Adult leaf
Roots
Root tip
Elongation zone
Lateral root
100
0
% of expression potential
Current Opinion in Plant Biology

Pollen tube reception and fungal invasion recruit the same *FER*-dependent pathway

NTA is a member of the plant-specific MILDEW RESISTANCE LOCUS O (MLO) family of proteins (*45*). The MLO protein family was first identified because it plays a role in powdery mildew resistance and mutations in *MLO* genes confer resistance to powdery mildew, while presence of the wild-type proteins causes susceptibility (*50,51*). Fungal invasion is analogous to the fertilization process and strongly relies on cell-cell communication: when a fungal spore lands on an epidermal cell, it hydrates, germinates, and the tip-growing fungal hyphae penetrate the host cell wall in order to establish feeding structures called haustoria (*52*). The host cell reacts with re-organization of the cell polarity, which includes the transport of MLO proteins towards the penetration site (*53*) and will eventually lead to a reinforcement of the cell wall (*52*). NTA is localized throughout the cytoplasm of the synergids before fertilization, and gets polarly localized to the basal part of the synergid cells after arrival of the pollen tube (*45*). However, this does not occur in *fer* mutant embryo sacs, indicating that the re-localization of NTA fully depends on a functional *FER* signaling pathway. Supporting this link between fungal invasion and pollen tube reception, *fer* homozygous mutants are resistant to infection by the powdery mildew *Golovinomyces (syn. Erysiphe) orontii*, which infects *Arabidopsis* (Figure 1a) (*45*). This suggests that the correct positioning of NTA or other MLO proteins to the filiform apparatus and the fungal penetration site, respectively, are crucial for successful pollen tube reception and fungal invasion. Their re-localization depends on *FER*, which might recognize the cell wall changes caused by either invading pollen tubes or fungal hyphae.

Figure 1: (**a**) Phenotypes of CrRLK1L subfamily mutants during plant development and reproduction. Explanation in clockwise rotation starting at the flowers: Aniline-blue staining of callose in pollen tubes two days after pollination. Fertilized wild-type ovule (left) and *fer* mutant embryo sac (right) with abnormal pollen tube reception (pollen tube overgrowth). In *anx1/anx2* double mutants (right) pollen tubes burst directly after germination *in vitro*, whereas wild-type pollen tubes (left) are rapidly growing. During vegetative growth *the1/herk1* double mutants show a severe phenotype (short petioles), which is comparable to the dwarf phenotype of homozygous *fer-1* and *fer-2* mutants. Powdery mildew infected plants with fungal mycelium and conidiophores on the wild-type leaf (left), whereas the *fer-1* mutant is resistant (right). Root hairs of 4-day old *fer-4* mutant seedlings display a severe defect with short, collapsing, and bursting root hairs, compared to the wild type (left). The *the1* mutant was found as suppressor of the cellulose synthase-deficient mutant *prc1*, partially rescuing its short-hypocotyl phenotype in 5-day old dark-grown seedlings (right). Dark-grown etiolated *fer-2* seedlings display an enhanced ethylene response after 4 days of growth in ethylene, leading to shorter hypocotyls in the mutant (middle) compared to the wild type (left) and *fer-2* mutants complemented with a FER-GFP construct (right). Figure 1 (a) is adapted from (19,24,25,45) (**b**) Relative gene expression of *CrRLK1Ls* throughout the plant according to Genevestigator microarray database using Meta-Profile Analysis tool, Anatomy Profile (76).

Biotic and abiotic stress responses modulate CrRLK1L transcript levels

As abiotic stresses and microorganisms can deeply impact plant growth and the molecular and mechanical properties of the cell wall, it is not surprising that plants respond to these extracellular challenges by modulating the expression levels of the *CrRLK1L* gene family. Among them, *THE1* is the member whose transcript levels are influenced the most by abiotic stresses, while *At5g39030* transcript levels do not show dramatic changes (Figure 2). The global plant response to environmental stresses is complex because plants integrate the environmental challenges at every organization level in both time and space. Abiotic stresses trigger various and numerous effects in plants, and eventually lead to growth inhibition. Besides ion imbalance, one of the consensus stress-induced effects is the enhanced accumulation of ROS, which can become harmful to plant cells by damaging lipids, proteins, DNA, and carbohydrates (*e.g.* (*54*)). Therefore, to cope efficiently with surrounding stresses, plants have to shift the balance of ROS detoxification/production towards detoxification. Curiously, little variation in gene expression is observed for the ROS-producing enzymes, the Rboh NADPH-oxidases, under abiotic stresses (*55*). However, abiotic stresses result globally in down-regulation of *CrRLK1Ls*, with heat treatment and hypoxia being the most effective (Figure 2). Since an emerging role for the *CrRLK1L* subfamily appears to be the regulation of NADPH-oxidase-dependent ROS production (*20*,*25*), down-regulation of *CrRLK1L* transcript levels in response to abiotic stress could be a strategy plants have developed to reduce production of ROS. Although no abiotic stress-related phenotype has been described for any *CrRLK1L* mutants, it would be interesting to test whether some of these mutants, *e.g. the1* or *fer,* exhibit impaired response to abiotic stresses such as heat-shock protein accumulation under heat stress (*56*).

The functional importance among *CrRLK1Ls* during plant-microorganism interactions has so far only been revealed for *FER*, which mediates susceptibility to *E. orontii* and is affected in FLG22-induced ROS production, MAPK activity, and stomatal closure (*45,57*). To protect themselves from microorganisms, plants are able to sense danger by recognizing conserved pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) through pattern-recognition receptors (PRRs). They also recognize their own molecular patterns, referred to as damage-associated molecular patterns (DAMPs), which appear only when plant tissue is infected or damaged. Those molecular patterns are collectively called elicitors when applied externally. They rapidly induce plant innate

immune responses characterized by a quick increase in cytosolic Ca^{2+}, MAPK signaling activation, oxidative burst, cell wall reinforcements, and defense-related gene expression (*58*). Except for *THE1* and *At5g24010*, elicitor treatments result in a rapid increase of *CrRLK1L* transcript levels with *At5g39020* and *At5g38990* being up-regulated the most consistently across elicitor treatments, followed by *At3g39030*, *At2g39360* and *At2g23200* (Figure 2). Moreover, the transcript levels of these five *CrRLK1Ls* are highly down-regulated in response to inoculation with virulent *Pseudomonas syringae pv. tomato* and this down-regulation is completely dependent on the ability of the bacteria to secrete effectors into the plant (Figure 2, see difference with *P. syringae pv. tomato* DC3000 *hrpA* mutant strain vs virulent DC3000 strain).

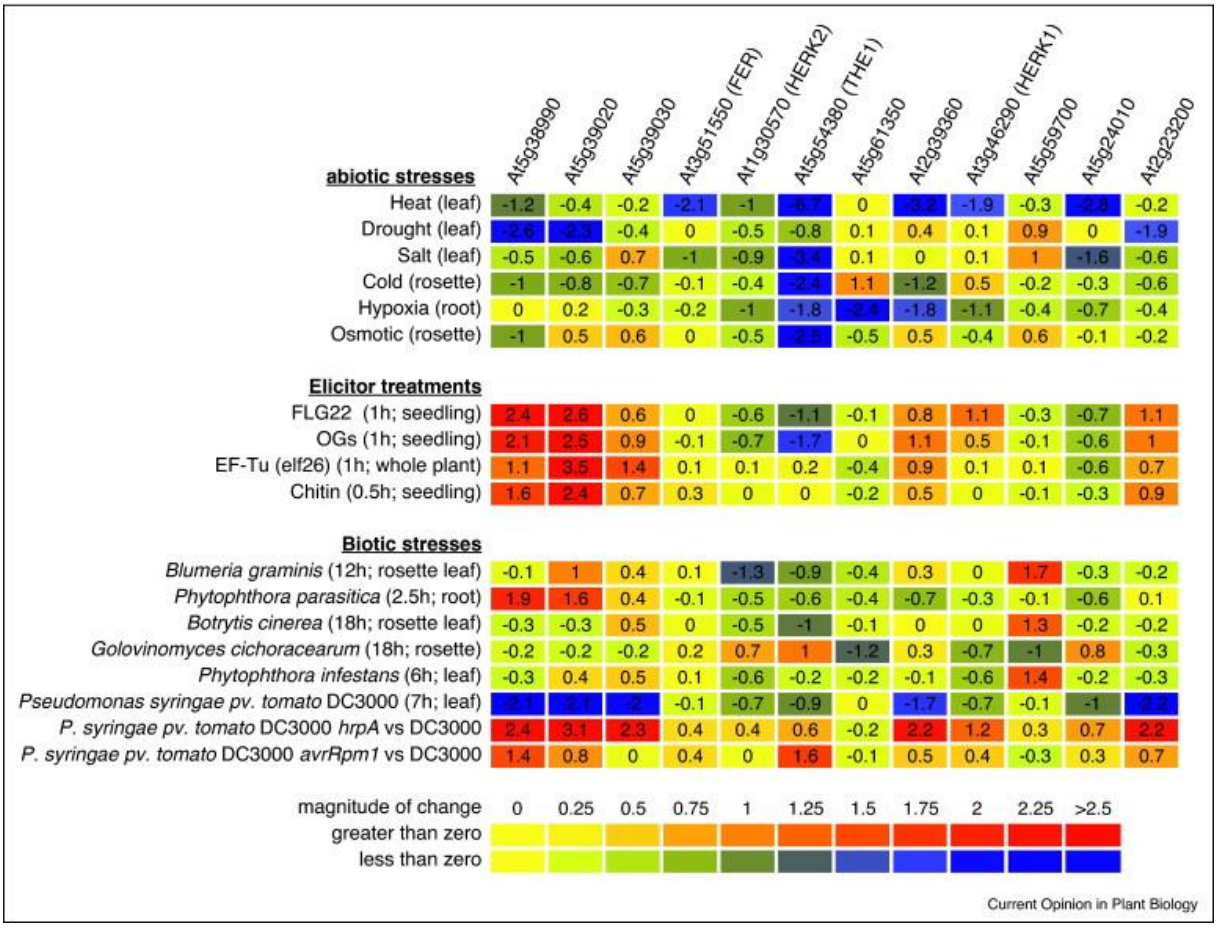


Figure 2: Significant transcript level changes for the *Arabidopsis CrRLK1L* gene subfamily under abiotic, elicitor, and biotic treatments. Changes in steady-state transcript levels for the *CrRLK1Ls* that are not preferentially expressed in pollen were obtained from the Genevestigator microarray database (76). Data is represented as log(2)-ratio versus mock or untreated control (if not specified), and is filtered via the Genevestigator Perturbations tool for treatments that induce a fold-change >2 (p-value <0.05) for at least one member of the *CrRLK1Ls.* The heat map was generated using the HeatMapper Tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper.cgi). Note that *At5g39000* is represented by the same probe as *At5g38990*.

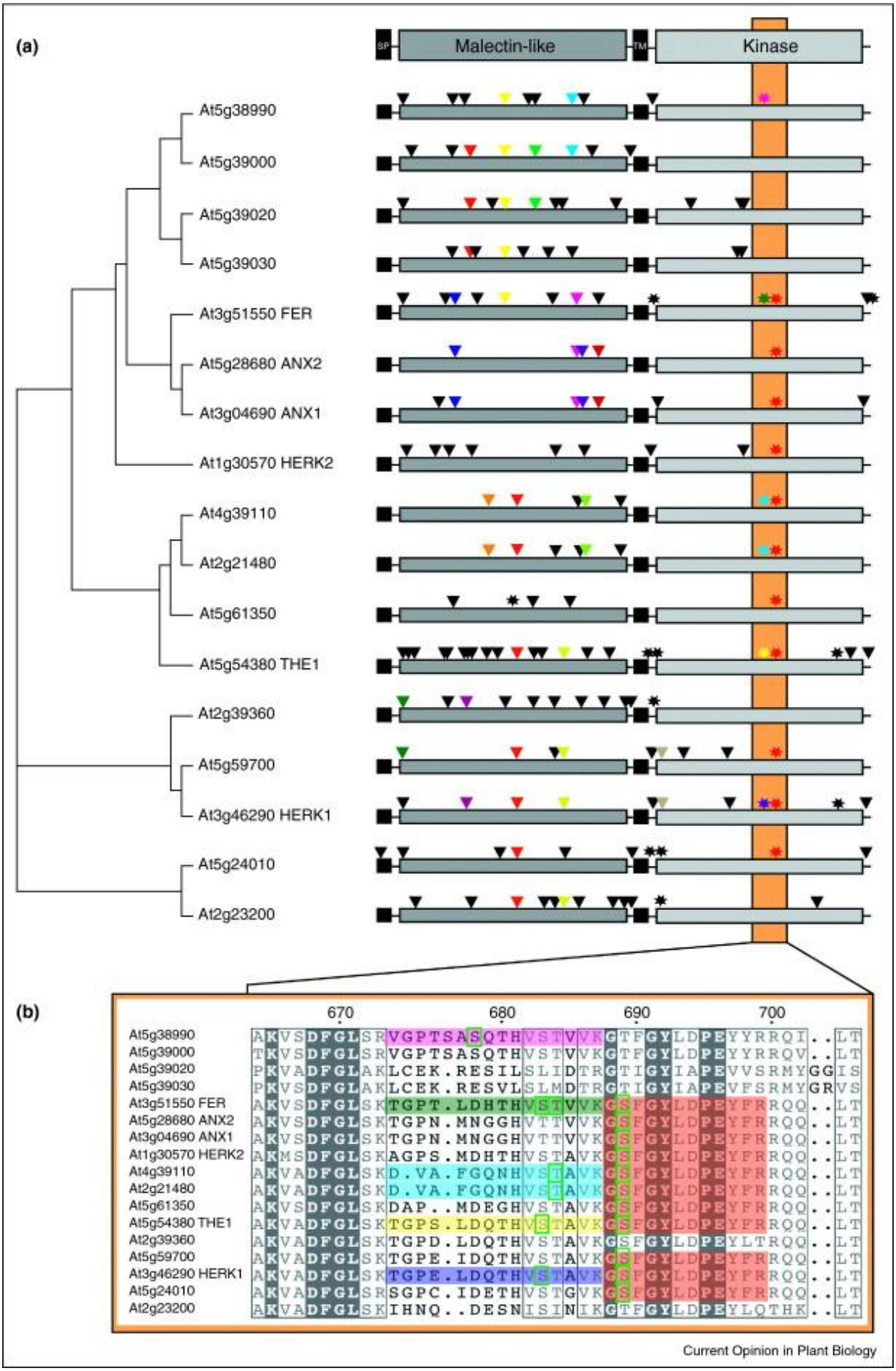
As secreted bacterial effectors suppress PAMP-triggered responses (*59*), it would be worth to investigate the role of those *CrRLK1Ls* in PAMP-triggered immunity; for example they could be part of and regulate PAMP RLK complexes. In this respect, the fact that the *P. syringae* effector AvrPto interacts *in vitro* with and inhibits the autophosphorylation of the elicitor-induced CrRLK1L At2g23200 as well as the PRRs FLAGELLLIN-SENSING2 (FLS2) and the EF-Tu receptor (EFR) could be of significant importance (*60*).

POST-TRANSLATIONAL MODIFICATIONS COULD BE CRUCIAL FOR *CrRLK1L* FUNCTION

Phosphorylation of the kinase domain might be involved in signal transduction

Protein regulation by phosphorylation or dephosphorylation is an important means to modulate cellular processes and is accomplished by protein kinases and phosphatases, respectively. It has been shown *in vitro* that the kinase domains of FER and At2g23200 are active and able to autophosphorylate (*17,60*); however, the biological relevance of phosphorylation is unknown. Several studies have experimentally assessed the phosphoproteome of various plant tissues under different environmental conditions, discovering several phosphopeptides that could be assigned to various members of the CrRLK1L subfamily (Figure 3) (*61-65*).

Figure 3: Phylogenetic tree, domain organization, and post-translational modifications of the CrRLK1Ls. (**a**) Phylogenetic tree of the 17 members of the *Arabidopsis* CrRLK1L subfamily with their domain organization and post-translational modifications. Triangles: putative glycosylation sites based on prediction by NetNGlyc1.0. Only significant sites are depicted. Black triangles: unique glycosylation sites. Colorful triangles: glycosylation sites that are conserved in their location and sequence between family members. Asterisks: Phosphorylation sites. Black asterisks: unique phosphosites. Red asterisks: conserved phosphosite between 11 members of the CrRLK1L subfamily that lies within the activation loop of the kinase domain. Colorful asterisks: phosphosites that are within the activation loop, but not conserved between members or only between two close relatives (blue asterisks). The colors of the asterisks are consistent with the color code in panel (b). (**b**) Multiple sequence alignment of a section of the activation loop of the kinase domain of CrRLK1Ls. Shaded in red: phosphopeptide that is conserved between 11 members. Shaded in colors: phosphopeptides within the activation loop that are not conserved or only conserved between two members (blue). Note that the color code is consistent with the asterisks in panel (a). Green rectangles: experimentally validated/confirmed phosphorylation sites. Phosphorylation sites and phosphopeptides in panels (a) and (b) are based on (61-66,77).


(a)
SP
Malectin-like
TM
Kinase
At5g38990
At5g39000
At5g39020
At5g39030
At3g51550 FER
At5g28680 ANX2
At3g04690 ANX1
At1g30570 HERK2
At4g39110
At2g21480
At5g61350
At5g54380 THE1
At2g39360
At5g59700
At3g46290 HERK1
At5g24010
At2g23200
(b)
670
680
690
700
At5g38990 AKVSDFGLSRVGPTSASQTHVSTVVKGTFGYLDPEYYRRQI..LT
At5g39000 TKVSDFGLSRVGPTSASQTHVSTVVKGTFGYLDPEYYRRQV..LT
At5g39020 PKVADFGLAKLCEK.RESILSLIDTRGTIGYIAPEVVSRMYGGIS
At5g39030 PKVSDFGLAKLCEK.RESVLSLMDTRGTIGYIAPEVFSRMYGRVS
At3g51550 FER AKVSDFGLSKTGPT.LDHTHVSTVVKGSFGYLDPEYFRRQQ..LT
At5g28680 ANX2 AKVSDFGLSKTGPN.MNGGHVTTVVKGSFGYLDPEYFRRQQ..LT
At3g04690 ANX1 AKVSDFGLSKTGPN.MNGGHVTTVVKGSFGYLDPEYFRRQQ..LT
At1g30570 HERK2 AKMSDFGLSKAGPS.MDHTHVSTAVKGSFGYLDPEYFRRQQ..LT
At4g39110 AKVADFGLSKD.VA.FGQNHVSTAVKGSFGYLDPEYFRRQQ..LT
At2g21480 AKVADFGLSKD.VA.FGQNHVSTAVKGSFGYLDPEYFRRQQ..LT
At5g61350 AKVSDFGLSKDAP..MDEGHVSTAVKGSFGYLDPEYFRRQQ..LT
At5g54380 THE1 AKVADFGLSKTGPS.LDQTHVSTAVKGSFGYLDPEYFRRQQ..LT
At2g39360 AKVADFGLSKTGPD.LDQTHVSTAVKGSFGYLDPEYLTRQQ..LT
At5g59700 AKVADFGLSKTGPE.IDQTHVSTAVKGSFGYLDPEYFRRQQ..LT
At3g46290 HERK1 AKVADFGLSKTGPE.LDQTHVSTAVKGSFGYLDPEYFRRQQ..LT
At5g24010 AKVADFGLSRSGPC.IDETHVSTGVKGSFGYLDPEYFRRQQ..LT
At2g23200 AKVADFGLSKIHNQ..DESNISINIKGTFGYLDPEYLQTHK..LT
Current Opinion in Plant Biology

Only recently, a study investigating the pollen phosphoproteome was published (*66*). This is especially interesting because *de novo* transcription is not essential for pollen germination and initial tube growth. Therefore, the cellular machinery leading to pollen germination and rapid tip growth must be regulated at the post-transcriptional level. In their study, Mayank and coworkers identified two different ambiguous phosphopeptides representing the CrRLK1L subfamily (*66*). The first phosphopeptide, with a phosphorylated threonine, is only conserved between At2g21480 and At4g39110, whose genes are specifically expressed in pollen (Figure 1) (*15*). The second identified phosphopeptide, with a phosphorylated serine, perfectly matches 11 members of the CrRLK1L subfamily, including THE1, HERK1 and FER, but can be assigned to the four representatives that are preferentially expressed in pollen: ANX1 and ANX2 and the above-mentioned At2g21480 and At4g39110 (Figure 3b).

Interestingly, the very same phosphopeptide has been identified before in suspension-cultured cells and lies within the activation loop of the kinase domain of the CrRLK1Ls (*63*), pointing towards the importance of an active kinase domain in all these proteins. It will be interesting to experimentally verify the biological importance of the phosphorylation sites *in vivo,* for example by direct manipulation of the relevant amino acids and mutant complementation assays. So far the only known interactors of the intracellular domain for CrRLK1L members are ROPGEFs, ROPs and AvrPto (*25,60*). Further studies addressing whether the CrRLK1Ls are able to directly phosphorylate ROPGEFs/ROPs or if other kinases are involved are awaited for.

Ligand binding may depend on N-linked glycosylation of the extracellular domain

Recently, several studies have proposed a role in pollen tube guidance and fertilization for proteins involved in endoplasmic reticulum (ER) quality control of secreted proteins (*67,68*) and glycosylation (*49*), respectively. Whereas the pollen tubes of *pollen defective in guidance 1* (*pod1*) and *cation/proton exchanger chx21/chx23* mutants are impaired in targeting embryo sacs, *tun* mutant female gametophytes are defective in pollen tube reception. *POD1* encodes a protein of unknown function, and localizes, like CHX23, to the ER. *TUN* encodes a UDP-Glycosyltransferase superfamily protein, which might play a role in glycosylating proteins in the ER. N-linked glycosylation refers to the co- and post-translational covalent attachment of an oligosaccharide (Glc_3-Man_9-$GlcNAc_2$ in eukaryotes) in the N-X-S/T context, where X stands for every amino acid except proline

(*69*). After attachment of the oligosaccharide (glycan) to a protein in the ER, it can be further modified in the secretory pathway. In plants, the glycans on glycoproteins are important for the protein folding process within the ER (*70*). Beyond the ER, N-glycans affect catalytic activity, stability, and folding, as well as the subcellular localization and secretion of proteins. Furthermore, N-glycosylation seem to be involved in plant pathogen interactions and in the functional binding of ligands to PRRs belonging to the RLK family (*71*).

The extracellular region of the 17 members of the *Arabidopsis* CrRLK1L subfamily consists of two domains with limited homology to the carbohydrate-binding domain of Malectin, and is referred as Malectin-like domain (Figure 3) (*13*). Malectin was found as an ER-membrane localized lectin in *Xenopus laevis* and binds carbohydrates with selectivity for di-glucose-high mannose N-glycan (Glc_2-N-glycan) (*16*). More recently, this protein has been found to participate in mammalian ER-quality control for glycoproteins (*72*). Although CrRLK1L subfamily members contain a malectin-like domain, they are not thought to be involved in N-glycosylation due to their plasma membrane localization and the weak conservation of residues mediating the interaction with glucose residues in the original Malectin. Nevertheless, it is interesting to note that all 17 members have several putative N-glycosylation sites (3-16 sites), 43%-100% of those lie in the extracellular malectin-like domain (Figure 3a). Proteins with higher identities share several N-glycosylation sites even in the extracellular region, which is less conserved among the CrRLK1L members. The functional importance of the N-glycosylation sites in malectin-like domains is unknown, but Haweker and colleagues (*73*) showed that a mutation in a single conserved glycosylation site in the ECD of the EFR–RLK, involved in receptor-mediated immune response, leads to a loss of ligand (elf18) binding ability.

That glycosylation sites can be involved in ligand binding is further encouraged by the role they play during fertilization in animals. There, cross-linked glycoprotein filaments in the extracellular coat of the oocyte, the zona pellucida, mediate species-restricted recognition between the gametes (*74*). A mutation in a single glycosylation site of the zona pellucida glycoprotein 3 (ZP3) reduced sperm binding by 80% (*75*). One of the known roles of *FER* is to prepare the female gametophyte for fertilization. It is very likely that the ECD of this RLK interacts with signals outside the cell, which are transduced by a signaling cascade inside the synergid cell. The presence of a malectin-

like domain in its ECD leads us to speculate that the putative ligand of the CrRLK1L kinases could contain polysaccharides. Whether these putative ligands originate from degraded cell wall components or from secreted glycosylated proteins, and if some glycosylation sites in the malectin-like domain mediate ligand-binding, requires further investigations.

Conclusion

The plant cell wall has to be able to react quickly to external and internal signals. External signals can be abiotic stresses, biotic stresses such as pathogens, or invading pollen tubes. In contrast, internal signals originate from the plant's own tissues, leading to adaptations such as cell elongation during plant growth. Characterized members of the CrRLK1L subfamily seem to play a crucial role in sensing cell wall perturbations. Signal perception and transduction could rely on post-translational modifications such as N-glycosylation and phosphorylation. These modifications can be specific for different cell types, tissues, and species. This could explain for example why *FER*, which is widely expressed throughout the plant, can accomplish diverse functions depending on the tissue or organ it is expressed in. We anticipate that further studies on this subfamily of RLKs will impact many important aspects of plant biology, including cell wall remodeling, glycobiology, cell growth, cell-cell communication, as well as plant responses to abiotic and biotic stresses.

Acknowledgements

We thank Cyril Zipfel, Robert Dudler and the members of the Grossniklaus group, in particular Sharon Kessler, for enriching discussions, and our esteemed colleagues Kian Hématy, Herman Höfte, Alice Cheung, Yanhai Yin and Paul Larsen for generously providing the original pictures shown in Figure 1. The work on the CrRLK1L family in our lab is supported by the University of Zürich, the *Forschungskredit* der *Universität Zürich* and the European Union through a Marie Curie International Reintegration grant to A.B.-D., grants of the Swiss National Science Foundation (31003A-112489) and SystemsX.ch ('Plant Growth') to U.G., and support for H.L. and L.M.M. through Research Modules of the ProDoc Programs 'Molecular Life Sciences' and 'Plant Science and Policy' to U.G.

References

1. Popper ZA, Michel G, Herve C, Domozych DS, Willats WG, Tuohy MG, Kloareg B, Stengel DB: **Evolution and diversity of plant cell walls: from algae to flowering plants**. *Annu Rev Plant Biol* 2011, **62**:567-590.

2. Cosgrove DJ: **Growth of the plant cell wall**. *Nat Rev Mol Cell Biol* 2005, **6**:850-861.

3. Wolf S, Hematy K, Hofte H: **Growth control and cell wall signaling in plants**. *Annu Rev Plant Biol* 2012, **63**:381-407.

4. Levin DE: **Regulation of cell wall biogenesis in *Saccharomyces cerevisiae*: the cell wall integrity signaling pathway**. *Genetics* 2011, **189**:1145-1175.

5. Ernst JF, Pla J: **Signaling the glycoshield: maintenance of the *Candida albicans* cell wall**. *Int J Med Microbiol* 2011, **301**:378-383.

6. Ellis C, Karafyllidis I, Wasternack C, Turner JG: **The *Arabidopsis* mutant *cev1* links cell wall signaling to jasmonate and ethylene responses**. *Plant Cell* 2002, **14**:1557-1566.

7. Cano-Delgado A, Penfield S, Smith C, Catley M, Bevan M: **Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana***. *Plant J* 2003, **34**:351-362.

8. Shiu SH, Bleecker AB: **Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis***. *Plant Physiol* 2003, **132**:530-543.

9. Gish LA, Clark SE: **The RLK/Pelle family of kinases**. *Plant J* 2011, **66**:117-127.

10. Steinwand BJ, Kieber JJ: **The role of receptor-like kinases in regulating cell wall function**. *Plant Physiol* 2010, **153**:479-484.

11. Seifert GJ, Blaukopf C: **Irritable walls: the plant extracellular matrix and signaling**. *Plant Physiol* 2010, **153**:467-478.

12. Ringli C: **Monitoring the outside: cell wall-sensing mechanisms**. *Plant Physiol* 2010, **153**:1445-1452.

13. Boisson-Dernier A, Kessler SA, Grossniklaus U: **The walls have ears: the role of plant CrRLK1Ls in sensing and transducing extracellular signals**. *J Exp Bot* 2011, **62**:1581-1591.

14. Cheung AY, Wu HM: **THESEUS 1, FERONIA and relatives: a family of cell wall-sensing receptor kinases?** *Curr Opin Plant Biol* 2011, **14**:632-641.

15. Hematy K, Hofte H: **Novel receptor kinases involved in growth regulation**. *Curr Opin Plant Biol* 2008, **11**:321-328.

16. Schallus T, Jaeckh C, Feher K, Palma AS, Liu Y, Simpson JC, Mackeen M, Stier G, Gibson TJ, Feizi T, et al.: **Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation**. *Mol Biol Cell* 2008, **19**:3404-3414.

17. Escobar-Restrepo JM, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang WC, Grossniklaus U: **The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception**. *Science* 2007, **317**:656-660.

18. Hematy K, Sado PE, Van Tuinen A, Rochange S, Desnos T, Balzergue S, Pelletier S, Renou JP, Hofte H: **A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis**. *Curr Biol* 2007, **17**:922-931.

19. Guo H, Li L, Ye H, Yu X, Algreen A, Yin Y: **Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana***. *Proc Natl Acad Sci U S A* 2009, **106**:7648-7653.

20. Denness L, McKenna JF, Segonzac C, Wormit A, Madhou P, Bennett M, Mansfield J, Zipfel C, Hamann T: **Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in *Arabidopsis***. *Plant Physiol* 2011, **156**:1364-1374.

21. Nemhauser JL, Mockler TC, Chory J: **Interdependency of brassinosteroid and auxin signaling in *Arabidopsis*.** *PLoS Biol* 2004, **2**:E258.

22. Kim TW, Wang ZY: **Brassinosteroid signal transduction from receptor kinases to transcription factors**. *Annu Rev Plant Biol* 2010, **61**:681-704.

23. Guo H, Ye H, Li L, Yin Y: **A family of receptor-like kinases are regulated by BES1 and involved in plant growth in *Arabidopsis thaliana***. *Plant Signal Behav* 2009, **4**:784-786.

24. Deslauriers SD, Larsen PB: ***FERONIA* is a key modulator of brassinosteroid and ethylene responsiveness in *Arabidopsis* hypocotyls**. *Mol Plant* 2010, **3**:626-640.

25. Duan Q, Kita D, Li C, Cheung AY, Wu HM: **FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development**. *Proc Natl Acad Sci U S A* 2010, **107**:17821-17826.

26. Gu Y, Li S, Lord EM, Yang Z: **Members of a novel class of *Arabidopsis* Rho guanine nucleotide exchange factors control Rho GTPase-dependent polar growth**. *Plant Cell* 2006, **18**:366-381.

27. Berken A, Thomas C, Wittinghofer A: **A new family of RhoGEFs activates the Rop molecular switch in plants**. *Nature* 2005, **436**:1176-1180.

28. Nibau C, Wu HM, Cheung AY: **RAC/ROP GTPases: 'hubs' for signal integration and diversification in plants**. *Trends Plant Sci* 2006, **11**:309-315.

29. Schiefelbein JW, Somerville C: **Genetic control of root hair development in *Arabidopsis thaliana***. *Plant Cell* 1990, **2**:235-243.

30. Monshausen GB, Bibikova TN, Messerli MA, Shi C, Gilroy S: **Oscillations in extracellular pH and reactive oxygen species modulate tip growth of *Arabidopsis* root hairs**. *Proc Natl Acad Sci U S A* 2007, **104**:20996-21001.

31. Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JD, et al.: **Reactive oxygen species produced by NADPH oxidase regulate plant cell growth**. *Nature* 2003, **422**:442-446.

32. Swanson S, Gilroy S: **ROS in plant development**. *Physiol Plant* 2010, **138**:384-392.

33. Boisson-Dernier A, Roy S, Kritsas K, Grobei MA, Jaciubek M, Schroeder JI, Grossniklaus U: **Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge**. *Development* 2009, **136**:3279-3288.

34. Miyazaki S, Murata T, Sakurai-Ozato N, Kubo M, Demura T, Fukuda H, Hasebe M: ***ANXUR1* and *2*, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization**. *Curr Biol* 2009, **19**:1327-1331.

35. Higashiyama T: **Peptide signaling in pollen-pistil interactions**. *Plant Cell Physiol* 2010, **51**:177-189.

36. Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D, et al.: **Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells**. *Nature* 2009, **458**:357-361.

37. Higashiyama T, Yabe S, Sasaki N, Nishimura Y, Miyagishima S, Kuroiwa H, Kuroiwa T: **Pollen tube attraction by the synergid cell**. *Science* 2001, **293**:1480-1483.

38. Marton ML, Cordts S, Broadhvest J, Dresselhaus T: **Micropylar pollen tube guidance by *egg apparatus 1* of maize**. *Science* 2005, **307**:573-576.

39. Chen YH, Li HJ, Shi DQ, Yuan L, Liu J, Sreenivasan R, Baskar R, Grossniklaus U, Yang WC: **The central cell plays a critical role in pollen tube guidance in *Arabidopsis***. *Plant Cell* 2007, **19**:3563-3577.

40. Yu F, Shi J, Zhou J, Gu J, Chen Q, Li J, Cheng W, Mao D, Tian L, Buchanan BB, et al.: **ANK6, a mitochondrial ankyrin repeat protein, is required for male-female gamete recognition in *Arabidopsis thaliana***. *Proc Natl Acad Sci U S A* 2010, **107**:22332-22337.

41. Huck N, Moore JM, Federer M, Grossniklaus U: **The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception**. *Development* 2003, **130**:2149-2159.

42. Rotman N, Rozier F, Boavida L, Dumas C, Berger F, Faure JE: **Female control of male gamete delivery during fertilization in *Arabidopsis thaliana***. *Curr Biol* 2003, **13**:432-436.

43. Huang BQaR, S.D.: **Female germ unit: organization, isolation, and function**. *Int. Rev. Cytol.* 1992, **140**.

44. Kanaoka MM, Torii KU: **FERONIA as an upstream receptor kinase for polar cell growth in plants**. *Proc Natl Acad Sci U S A* 2010, **107**:17461-17462.

45. Kessler SA, Shimosato-Asano H, Keinath NF, Wuest SE, Ingram G, Panstruga R, Grossniklaus U: **Conserved molecular components for pollen tube reception and fungal invasion**. *Science* 2010, **330**:968-971.

46. Capron A, Gourgues M, Neiva LS, Faure JE, Berger F, Pagnussat G, Krishnan A, Alvarez-Mejia C, Vielle-Calzada JP, Lee YR, et al.: **Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene**. *Plant Cell* 2008, **20**:3038-3049.

47. Rotman N, Gourgues M, Guitton AE, Faure JE, Berger F: **A dialogue between the *SIRENE* pathway in synergids and the fertilization independent seed pathway in the central cell controls male gamete release during double fertilization in Arabidopsis**. *Mol Plant* 2008, **1**:659-666.

48. Boisson-Dernier A, Frietsch S, Kim TH, Dizon MB, Schroeder JI: **The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition**. *Curr Biol* 2008, **18**:63-68.

49. Lindner H, Raissig MT, Sailer C, Shimosato-Asano H, Bruggmann R, Grossniklaus U: **SNP-Ratio Mapping (SRM): Identifying lethal alleles and mutations in complex genetic backgrounds by Next-Generation Sequencing**. *Genetics* 2012, DOI: 10.1534/genetics.112.141341.

50. Consonni C, Humphry ME, Hartmann HA, Livaja M, Durner J, Westphal L, Vogel J, Lipka V, Kemmerling B, Schulze-Lefert P, et al.: **Conserved requirement for a plant host cell protein in powdery mildew pathogenesis**. *Nat Genet* 2006, **38**:716-720.

51. Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, et al.: **The barley *Mlo* gene: a novel control element of plant pathogen resistance**. *Cell* 1997, **88**:695-705.

52. O'Connell RJ, Panstruga R: **Tête-a-tête inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes**. *New Phytol* 2006, **171**:699-718.

53. Bhat RA, Miklis M, Schmelzer E, Schulze-Lefert P, Panstruga R: **Recruitment and interaction dynamics of plant penetration resistance components in a plasma membrane microdomain**. *Proc Natl Acad Sci U S A* 2005, **102**:3135-3140.

54. Pucciariello C, Banti V, Perata P: **ROS signaling as common element in low oxygen and heat stresses**. *Plant Physiol Biochem* 2012, DOI: 10.1016/j.plaphy.2012.02.016.

55. Suzuki N, Miller G, Morales J, Shulaev V, Torres MA, Mittler R: **Respiratory burst oxidases: the engines of ROS signaling**. *Curr Opin Plant Biol* 2011, **14**:691-699.

56. Saidi Y, Finka A, Goloubinoff P: **Heat perception and signalling in plants: a tortuous path to thermotolerance**. *New Phytol* 2011, **190**:556-565.

57. Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Grossniklaus U, Schulze WX, Robatzek S, Panstruga R: **PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity**. *J Biol Chem* 2010, **285**:39140-39149.

58. Segonzac C, Zipfel C: **Activation of plant pattern-recognition receptors by bacteria**. *Curr Opin Microbiol* 2011, **14**:54-61.

59. Block A, Alfano JR: **Plant targets for Pseudomonas syringae type III effectors: virulence targets or guarded decoys?** *Curr Opin Microbiol* 2011, **14**:39-46.

60. Xiang T, Zong N, Zou Y, Wu Y, Zhang J, Xing W, Li Y, Tang X, Zhu L, Chai J, et al.: ***Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases**. *Curr Biol* 2008, **18**:74-80.

61. Benschop JJ, Mohammed S, O'Flaherty M, Heck AJ, Slijper M, Menke FL: **Quantitative phosphoproteomics of**

early elicitor signaling in *Arabidopsis*. *Mol Cell Proteomics* 2007, **6**:1198-1214.

62. Nühse TS, Bottrill AR, Jones AM, Peck SC: **Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses**. *Plant J* 2007, **51**:931-940.

63. Nühse TS, Stensballe A, Jensen ON, Peck SC: **Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database**. *Plant Cell* 2004, **16**:2394-2405.

64. Engelsberger WR, Schulze WX: **Nitrate and ammonium lead to distinct global dynamic phosphorylation patterns when resupplied to nitrogen-starved *Arabidopsis* seedlings**. *Plant J* 2012, **69**:978-995.

65. Whiteman SA, Serazetdinova L, Jones AM, Sanders D, Rathjen J, Peck SC, Maathuis FJ: **Identification of novel proteins and phosphorylation sites in a tonoplast enriched membrane fraction of *Arabidopsis thaliana***. *Proteomics* 2008, **8**:3536-3547.

66. Mayank P, Grossman J, Wuest S, Boisson-Dernier A, Roschitzki B, Nanni P, Nuhse T, Grossniklaus U: **Characterization of the phosphoproteome of mature *Arabidopsis* pollen**. *Plant J* 2012, DOI: 10.1111/j.1365-313X.2012.05061.x.

67. Li HJ, Xue Y, Jia DJ, Wang T, Hi DQ, Liu J, Cui F, Xie Q, Ye D, Yang WC: ***POD1* regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis***. *Plant Cell* 2011, **23**:3288-3302.

68. Lu Y, Chanroj S, Zulkifli L, Johnson MA, Uozumi N, Cheung A, Sze H: **Pollen tubes lacking a pair of K^+ transporters fail to target ovules in *Arabidopsis***. *Plant Cell* 2011, **23**:81-93.

69. Bause E: **Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes.** *Biochem. J* 1983, **209**:331-336.

70. Helenius A, Aebi M: **Roles of N-linked glycans in the endoplasmic reticulum**. *Annu Rev Biochem* 2004, **73**:1019-1049.

71. Song W, Henquet MG, Mentink RA, van Dijk AJ, Cordewener JH, Bosch D, America AH, van der Krol AR: **N-glycoproteomics in plants: perspectives and challenges**. *J Proteomics* 2011, **74**:1463-1474.

72. Galli C, Bernasconi R, Solda T, Calanca V, Molinari M: **Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER**. *PLoS One* 2011, **6**:e16304.

73. Haweker H, Rips S, Koiwa H, Salomon S, Saijo Y, Chinchilla D, Robatzek S, von Schaewen A: **Pattern recognition receptors require N-glycosylation to mediate plant immunity**. *J Biol Chem* 2010, **285**:4629-4636.

74. Monne M, Jovine L: **A structural view of egg coat architecture and function in fertilization**. *Biol Reprod* 2011, **85**:661-669.

75. Han L, Monne M, Okumura H, Schwend T, Cherry AL, Flot D, Matsuda T, Jovine L: **Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3**. *Cell* 2010, **143**:404-415.

76. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P: **Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes**. *Adv Bioinformatics* 2008, **2008**:420747.

77. Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y: **Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis***. *Mol Syst Biol* 2008, **4**:193.

Chapter 2 – Results

The Role of *FERONIA* in Interspecific Pollen Tube Reception

The Role of *FERONIA* in Interspecific Pollen Tube Reception

Introduction

FERONIA and its role in pollen tube reception

FERONIA (*FER*) is a member of the *Catharanthus roseus* receptor-like serine/threonine kinase1-like (*CrRLK1L*) family (see introduction of this thesis; (*1*)). FER is plasma membrane-localized and contains besides its transmembrane domain an intracellular kinase domain and an extracellular malectin-like carbohydrate binding domain (*2*). *FER* is expressed throughout the plant except in pollen, and in ovules the strongest FER-GFP signal has been observed at the micropylar end of the synergid cells, the filiform apparatus (*3*). *FER*, which is allelic to *SIRENE* (*SIR*), has first been identified in a screen for reproductive mutants, and only later its additional functions in cell elongation during vegetative development, root hair tip growth and powdery mildew resistance have been revealed (*4-8*). The reproductive phenotype of heterozygous *fer* mutant plants is characterized by around 50% unfertilized ovules per silique and homozygous mutants are extremely scarce (*4*). The reason for the high numbers of unfertilized ovules is that in *fer* mutant ovules, after arrival at the micropylar opening the pollen tube (PT) does not stop its growth and does not burst in order to release its sperm cells, but instead continues growing within the ovule. This characteristic phenotype is named "PT overgrowth" and leaves the embryo sacs unfertilized (*4*). It is important to note that *fer* is a female gametophytic mutation, indicating that PT reception is under control of the embryo sac. Thus, in *fer* mutants the communication between the male and the female gametophyte is impaired. Meanwhile, several other female gametophytic mutants displaying a *fer*-like PT overgrowth phenotype have been identified. These mutants are *lorelei* (*lre*), which is a mutant in a gene coding for a glycosylphosphatidylinositol (GPI)-anchored protein, *nortia* (*nta*), a mutant in a Mildew Resistance Locus O (MLO)-like gene and *turan* (*tun*), which disrupts an UDP-glycosyltransferase (*8-10*). In addition, mutants in the peroxin gene *ABSTINENCE BY MUTUAL CONSENT* (*AMC*) display PT overgrowth in their ovules, however only if both PT and embryo sac are *amc* (*11*). Due to their membrane localization, it is conceivable that FER, NTA and LRE could be directly involved in the recognition of extracellular signals derived from or caused by the

arriving PT, whereas TUN and EVN more likely play an indirect role through glycosylation of putative receptors and/or ligands crucial for PT reception.

Interspecific crosses resemble the *feronia* phenotype

If FER is an active signaling component involved in direct recognition of PT-derived molecules, evolutionary changes in FER or its ligand could lead to impaired PT reception and act as a hybridization barrier. Indeed, in 20 different interspecific and intergeneric crosses of the genus *Rhododendron* (Ericaceae), a PT overgrowth phenotype reminiscent of *fer* has been observed (*12*, *13*). Likewise, PTs also overgrow in interspecific crosses within the Brassicaceae, where *A. thaliana* siliques pollinated with *A. lyrata* pollen displayed about 50% of ovules with PT overgrowth (*3*). If *Cardamine flexuosa* was used as a pollen donor, only a small proportion of ovules attracted PTs, but of the ones that did, 70% were defective in PT reception and showed PT overgrowth. The extracellular domain of FER, which is the domain likely responsible for putative ligand binding, showed high Ka/Ks values when sequences from *A. thalinana*, *A. lyrata* and *C. flexuosa* were compared, which is indicative of rapid evolution (*3*). Based on these results, it was postulated that FER is directly involved in PT-derived ligand recognition and therefore a critical player in the establishment of a reproductive isolation barrier within the Brassicaceae. However, this hypothesis has not been experimentally verified.

Hypothesis and experimental design

If FER is directly involved in the binding of a putative PT-derived ligand, and if this receptor-ligand binding constitutes a reproductive barrier, it is conceivable that this barrier could be overcome by expressing the *A. lyrata* homolog of *FER* in a *fer* homozygous mutant background. These plants would then only carry the *A. lyrata* version of the receptor and if our hypothesis was true, their ovules should perform better in the recognition of *A. lyrata* PTs than plants carrying the *A. thaliana FER*, because in that case, the putative receptor-ligand pair is derived from the same species.

Results

FER (*A. lyrata*) localizes to the filiform apparatus of synergid cells

In order to test if *A. thaliana* plants expressing the *A. lyrata* homolog of *FER* – which is further denoted "*FER* (*Alyr*)" – can recognize *A. lyrata* PTs better than plants expressing the (endogenous) *FER* version of *A. thaliana, "FER* (*Athal*)", we amplified the coding sequence of *FER* from *A. lyrata* DNA and created a *FER (Alyr)-GFP* fusion under the control of the *FER* promoter from L*er*, Col-0 and *A. lyrata*, respectively. We chose those three promoters in order to make sure that at least one construct is expressed in *A. thaliana*. The L*er* promoter differs from the Col-0 promoter in a 689kb insertion. All constructs were transformed into *A. thaliana fer* heterozygous mutants (L*er* background).

Confocal microscopy analyses revealed that two days after emasculation, the ovules displayed strong GFP signal at the filiform apparatus of the synergid cells, similar to the reported localization of FER (Athal)-GFP (*3*). The filiform apparatus-localization was the same for all constructs with each of the three tested promoters (Fig. 1).

Thus, we concluded that *FER (Alyr)-GFP* is expressed in synergid cells and the corresponding protein is correctly targeted to the site of PT reception in the synergids cells. For the following experiments, we continued working only with the construct driven by the *A. lyrata FER*-promoter.

FER (*A.lyrata*) can complement the *A. thaliana fer* phenotype

FER (Alyr) and FER (Athal) differ in 24 amino acids, one of which is in the kinase domain, four in the transmembrane domain and the residual 19 amino acid exchanges are found in the extracellular domain. In order to test if *FER (Alyr)-GFP* can complement the *fer* mutant phenotype in intraspecific PT reception, we analyzed the seed set of selfed siliques of plants homozygous for both *fer* and the transgene. Plants expressing *FER (Alyr)-GFP* in the *fer* homozygous mutant background show a seed set comparable to wild-type (92,6% in the transgenic and 96,0% fertilized seeds in the wild-type, respectively) as well as to plants expressing *FER (Athal)-GFP* in the *fer* background 94,4%, Fig. 2A), indicating that *FER* (*Alyr*) can fully complement the *fer* mutant phenotype.

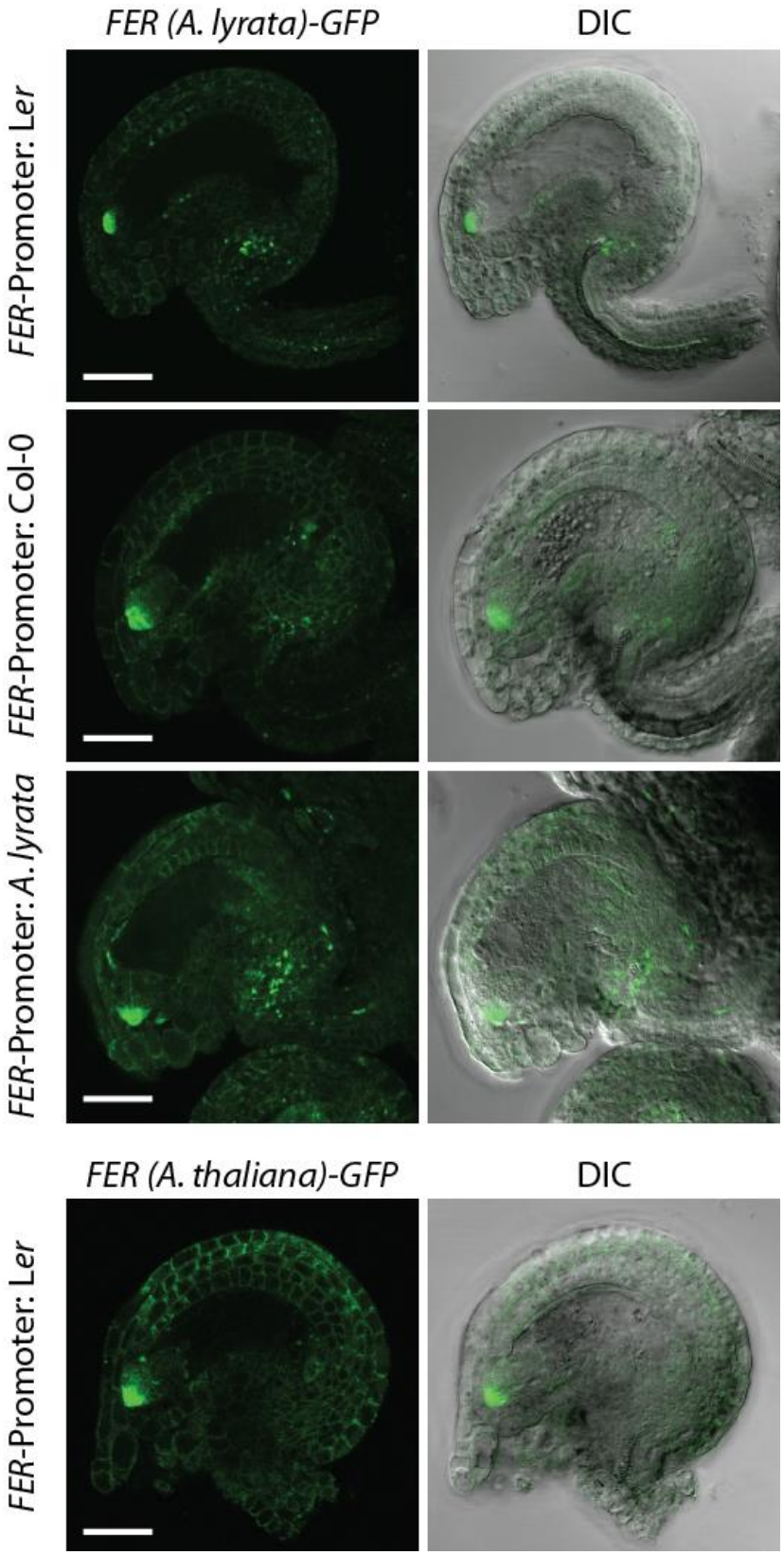


Figure 1: **Confocal images of ovules expressing *FER (Alyr)-GFP* driven by 3 different promoters.** FER (Alyr)-GFP localizes to the filiform apparatus of the synergid cells, exactly like it was observed for FER (Athal)-GFP (bottom panel). Scale bars: 20 µm.

FERONIA (*A. lyrata*) does not improve *A. lyrata* PT reception compared to *FERONIA* (*A. thaliana*)

If FER is directly involved in recognizing putative signals from the arriving PT in a species-specific manner, the *fer*-like PT overgrowth phenotype observed in *A. thaliana* pollinated with *A. lyrata* pollen should be complemented in plants expressing *FER (Alyr)-GFP*, but not in plants expressing *FER (Athal)-GFP*. We conducted interspecific crosses by pollinating wild-type and transgenic plants expressing either the *A. thaliana* or the *A. lyrata* variant of *FER* in the *fer* homozygous background with *A. lyrata* pollen. Then we assessed the amount of ovules with PT overgrowth with respect to all ovules in a silique that were targeted by a PT (denoted as "PT overgrowth per silique"). Although plants expressing *FER (Alyr)-GFP* had reduced levels of PT overgrowth per silique, also plants expressing *FER (Athal)-GFP* showed reduced levels of overgrowth compared to wild-type (Fig. 2B). This suggests that the decreased levels of PT overgrowth in *FER (Alyr)-GFP* expressing plants likely are caused by secondary effects of the transgene or by the *fer* background rather than by direct protein-protein interactions of FER (Alyr)-GFP and a PT-derived ligand.

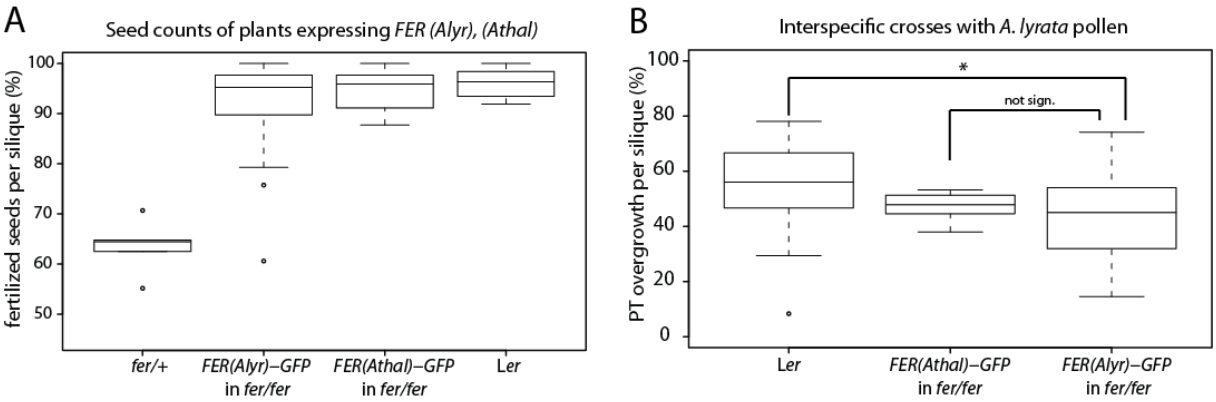


Figure 2: Seed count analysis of selfed plants and PT overgrowth in interspecific crosses of plants expressing *FER (Alyr)* or *FER (Athal)*. (**A**) Seed count analysis of plants expressing *FER (Alyr)-GFP* or *FER (Athal)-GFP* in the *fer* homozygous background. Both constructs have a seed set comparable to wild-type L*er* and do not differ significantly from it. The seed set of *fer* heterozygous plants (*fer*/+) is shown as a comparison. (**B**) Proportions of PT overgrowth per silique after interspecific crosses of wild-type and transgenic plants expressing *FER (Alyr)-GFP* or *FER (Athal)-GFP* in the *fer* homozygous background. Samples from both transgenic plants differ from L*er* with $p<0.05$ (*), but not from each other (chi-square test).

Discussion

Interspecific crosses between different members of the Brassicaceae family show a PT overgrowth phenotype that is reminiscent of the *fer* mutant. This and the fact that the extracellular domain of *FER* shows signs of rapid evolution, which is a typical feature of reproductive proteins, lead to the hypothesis that FER is directly involved in the recognition of putative ligands derived from the PT (*3*). However, our results show that expressing the *A. lyrata* homolog of *FER* in *A. thaliana* plants does not improve the reception of *A. lyrata* PTs, indicating that FER does not directly interact with putative PT ligands, at least not on the protein level. Of course it cannot be ruled out that other membrane-surface localized proteins of the FER-pathway, for example NTA or LRE, bind to possible PT ligands in a species-specific manner. The fact that both plants expressing *FER (Alyr)-GFP* as well was *FER (Athal)-GFP* seem to have reduced PT overgrowth in interspecific crosses compared to wild-type could be explained either by a secondary effect of the transgene itself or by effects of the *fer* homozygous mutant background. Indeed, most transgenic lines of either construct seemed to be smaller and weaker than the wild type and in addition seemed to have smaller pistils, possibly due to delayed development. Although all plants were pollinated two days after emasculation of the flower to ensure that all embryo sacs are in the same developmental stage, it is conceivable that immature ovules in those smaller pistils are more likely to receive an interspecific PT. This could be explained because potential barriers might not be established yet. A scenario like this would be similar to what was observed in interspecific incompatibilities in tomato and tobacco, where unilateral incompatibility barriers could be overcome in young pistils (*14*).

So far, the ligand of FER during reproduction has not been identified yet, but in roots, it has been shown that FER interacts with the rapid alkalization factor RALF1, a 5 kDa secreted peptide, and suppresses cell elongation (*15*). It was shown that RALF1-treatment of suspension cultures and seedlings decreases extracellular pH and rapidly increases cytoplasmic Ca^{2+} levels (*16*, *17*). There are 34 RALF-LIKE peptides in *A. thaliana*, and it is conceivable that one of them is the ligand of FER during PT reception. Initially, the CrRLK1L receptor kinases have been proposed to bind carbohydrates with their malectin-like domains in order to sense and regulate cell wall integrity (*2*). However, RALF1 does not contain any putative glycosylation sites and therefore might

bind at a site of the FER extracellular domain different from the malectin-like domains (*18*). Yet, it is possible that FER (and potentially also the other CrRLK1L proteins) do not only bind a single ligand but instead are receptors for multiple ligands that interact with different regions of the receptor protein.

Taken together, our results indicate that FER is most likely not directly involved in the recognition of interspecific PTs. Yet, it might still act as a receptor for PT-derived molecules, possibly in concert with another (co-)receptor which determines the species-specificity on the interaction. Further work is needed to identify the ligand(s) of FER during PT reception. This in turn might help to elucidate how interspecific crossing barriers during PT reception are established in the Brassicaceae. Furthermore, genetic screens aiming to identify novel factors involved in species-specific recognition of PTs will shed light on this process. Chapters 3 and 4 of this thesis present two new key players that mediate interspecific PT reception which were identified by exploiting natural variation in *A. thaliana*.

Material and Methods

Plant material and growth conditions

Plant growth conditions were as previously described (*8*). The *fer* mutant allele (*fer-1*) used for the described experiments is a *Ds* insertion line in the L*er* genomic background (*4*). The *A. lyrata* strain has been described in Escobar-Restrepo et al., 2007.

Vectors and plant transformation

The *FER (Athal)-GFP* construct has been described before (3). For *FER (Alyr)-GFP* under the control of the Col-0 and L*er* promoters, the respective promoters were amplified by using primers producing a 3' overhang that matches with the beginning of the *A. lyrata FER* coding sequence (the forward-primer introduces an attB1 site: 5' – GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTAAGCTTCGATTTAAGCG – 3' and the reverse primer overlaps with the coding sequence: 5' – GAGACGGAATTGTCCCTC – 3'). The *FER* coding sequence was amplified from *A. lyrata* genomic DNA with primers introducing a 5' overhang matching the Col-0 and L*er* promoter sequences, respectively (forward primer with overhang complementary to promoter sequence: 5' –

GAGGGACAATTCCGTCTC – 3' and the reverse primer introducing an attB2 site: 5' – GGGGACCACTTTGTACAAGAAAGCTGGGTACGTCCCTTTGGATTCATG – 3'). Then 50ng of each of the two PCR products were combined and used as a template in a third PCR using the forward and reverse primers with attB sites, thus pasting both products. The resulting PCR products were cloned into pDONR207 and subsequently into pMDC111 using Gateway Cloning ® (Life Technologies) following the instructions given by the manufacturer (*19*). *A. thaliana* plants were transformed using the floral dip method (*20*).

Confocal Analysis

Samples were prepared by dissecting pistils two days after emasculation in 1M Glycine (pH9.6 with KOH) and incubating the tissue on the slide for 24h at 4°C in the glycine solution for better clearing. Images were captured using the Leica SP2 Confocal Microscope.

Seed counts, crosses and Aniline Blue staining

Seed counts were performed by opening mature siliques and assessing the numbers of developing, plump seeds ("fertilized seeds") and unfertilized, white and shrunken ovules. For interspecific crosses, *A. thaliana* plants were emasculated and pollinated with *A. lyrata* pollen two days later. Two days after pollination, the elongated siliques were fixed for Aniline Blue staining, which was conducted as described previously (*4*).

References

1. H. Lindner, L. M. Muller, A. Boisson-Dernier, U. Grossniklaus, CrRLK1L receptor-like kinases: not just another brick in the wall, *Current opinion in plant biology* **15**, 659–669 (2012).

2. A. Boisson-Dernier, S. A. Kessler, U. Grossniklaus, The walls have ears: the role of plant CrRLK1Ls in sensing and transducing extracellular signals, *Journal of experimental botany* **62**, 1581–1591 (2011).

3. J. M. Escobar-Restrepo *et al.*, The *FERONIA* receptor-like kinase mediates male-female interactions during pollen tube reception, *Science* **317**, 656–660 (2007).

4. N. Huck, J. M. Moore, M. Federer, U. Grossniklaus, The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception, *Development* **130**, 2149–2159 (2003).

5. N. Rotman *et al.*, Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*, *Current biology* **13**, 432–436 (2003).

6. H. Guo, H. Ye, L. Li, Y. Yin, A family of receptor-like kinases are regulated by *BES1* and involved in plant growth in *Arabidopsis thaliana*, *Plant Signal Behav* **4**, 784–786 (2009).

7. Q. Duan, D. Kita, C. Li, A. Y. Cheung, H. M. Wu, *FERONIA* receptor-like kinase regulates RHO GTPase signaling of root hair development, *Proceedings of the National Academy of Sciences* **107**, 17821–17826 (2010).

8. S. A. Kessler *et al.*, Conserved molecular components for pollen tube reception and fungal invasion, *Science* **330**, 968–971 (2010).

9. A. Capron *et al.*, Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene, *The Plant cell* **20**, 3038–3049 (2008).

10. H. Lindner *et al.*, SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing, *Genetics* (2012).

11. A. Boisson-Dernier, S. Frietsch, T. H. Kim, M. B. Dizon, J. I. Schroeder, The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition, *Current biology* **18**, 63–68 (2008).

12. V. R. J. L. W. E. G. Kaul, Early events in the embryo sac after intraspecific and interspecific pollinations in *Rhododendron kawakamii* and *R. retusum*, *Canadian Journal of Botany* **64**, 282–291 (1986).

13. E. G. Williams, V. Kaul, J. L. Rouse, B. F. Palser, Overgrowth of Pollen Tubes in Embryo Sacs of *Rhododendron* Following Interspecific Pollinations, *Aust. J. Bot.* **34**, 413–423 (1986).

14. R. Swanson, A. F. Edlund, D. Preuss, Species Specificity In Pollen-Pistil Interactions, *Annu. Rev. Genet.* **38**, 793–818 (2004).

15. M. Haruta, G. Sabat, K. Stecker, B. B. Minkoff, M. R. Sussman, A peptide hormone and its receptor protein kinase regulate plant cell expansion, *Science* **343**, 408–411 (2014).

16. G. Pearce, D. S. Moura, J. Stratmann, C. A. Ryan, RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development, *Proceedings of the National Academy of Sciences* **98**, 12843–12847 (2001).

17. M. Haruta, G. Monshausen, S. Gilroy, M. R. Sussman, A cytoplasmic Ca2+ functional assay for identifying and purifying endogenous cell signaling peptides in *Arabidopsis* seedlings: identification of *At*RALF1 peptide, *Biochemistry* **47**, 6311–6321 (2008).

18. S. Wolf, H. Hofte, Growth Control: A Saga of Cell Walls, ROS, and Peptide Receptors, *The Plant cell* (2014).

19. M. D. Curtis, U. Grossniklaus, A gateway cloning vector set for high-throughput functional analysis of genes in planta, *Plant physiology* **133**, 462–469 (2003).

20. S. J. Clough, A. F. Bent, Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, *The Plant journal* **16**, 735–743 (1998).

Chapter 3 – Results

Interspecific Hybridization Barriers in Plants: GWAS Reveals a Role for Glycosylation in Gametophyte Recognition

This chapter has been submitted for publication in the present form.

Interspecific Hybridization Barriers in Plants: GWAS Reveals a Role for Glycosylation in Gametophyte Recognition

Lena M. Müller, Heike Lindner, Nuno D. Pires and Ueli Grossniklaus*

Institute of Plant Biology & Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

*Corresponding author. E-mail: grossnik@botinst.uzh.ch

Abstract

Species-specific gamete recognition is a key premise to ensure reproductive success and the maintenance of species boundaries. Because plant gametes interact only at the very last step of reproduction, it is likely that species-specific recognition occurs before that. At pollen tube (PT) reception, for instance, direct gametophyte interactions likely allow the recognition of species-specific signals from the PT by the embryo sac, resulting in PT rupture, sperm release, and double fertilization. This process is impaired in interspecific crosses between *Arabidopsis thaliana* and related species, leading to PT overgrowth and a failure to deliver the sperm cells to the female gametes. In a genome-wide association study (GWAS) for PT overgrowth in interspecific crosses between 86 *A. thaliana* accessions and *A. lyrata* pollen, we identified *ARTUMES (ARU)*, which specifically regulates the recognition of interspecific but not intraspecific PTs in *A. thaliana*. *ARU* encodes the OST3/6 subunit of the oligosaccharyltransferase complex conferring protein N-glycosylation. Our results suggest that glycosylation patterns of putative receptor-ligand pairs are crucial for gametophyte recognition in plants, and may represent an important mechanism of speciation.

Results and Discussion

Species evolve and are maintained by a variety of hybridization barriers that prevent interspecific gene flow. In contrast to pre-pollination barriers, which can be spatial or temporal patterns preventing plants from being pollinated by pollen of different species, post-pollination barriers come into play after pollination and can be divided into pre- and post-zygotic barriers (*1*). Pre-zygotic barriers prevent the fertilization of the female gametophyte (embryo sac) by the male gametophyte (pollen) and thus the formation of a zygote, while post-zygotic barriers are often caused by karyotype incompatibilities leading to hybrid inviability or sterility. The genic bases of such barriers constitute speciation genes that reduce gene flow between diverging lineages (*2*). While there are many studies on pre-pollination and post-zygotic hybridization barriers (reviewed in (*2*)), the molecular basis of post-pollination, pre-zygotic barriers is largely unknown. These barriers are established by pollen-pistil interactions and several studies describe a species-preferential behavior of molecular factors involved in pollen adherence to the stigma, pollen tube (PT) growth, and PT guidance towards the ovules (*3*-*8*). Importantly, all these factors act primarily in intraspecific pollination events and have additional species-preferential effects. Here, we describe the first gene that exclusively influences interspecific - but not intraspecific - gametophyte interactions and thus constitutes a molecular component of a hybridization barrier.

The self-fertilizing *A. thaliana* and its outcrossing relative *A. lyrata* are separated by strong pre-pollination barriers due to their different mating systems. In addition, they are isolated by post-pollination barriers based on direct male-female interactions. Whereas *A. thaliana* pollen germination is inhibited at the *A. lyrata* stigma, *A. lyrata* PTs can grow towards *A. thaliana* embryo sacs, but PT reception fails (fig. S1). This unilateral incompatibility is similar to that observed in other crosses between self- compatible and self-incompatible species (*9*). During PT reception, intercellular communication between male and female gametophytes (Fig. 3.1A) prepares the receptive synergid cell for penetration by the PT, which ruptures and releases the sperm cells to effect double fertilization (Fig. 3.1B). Thus, success or failure of PT reception is under female gametophytic control (*10*).

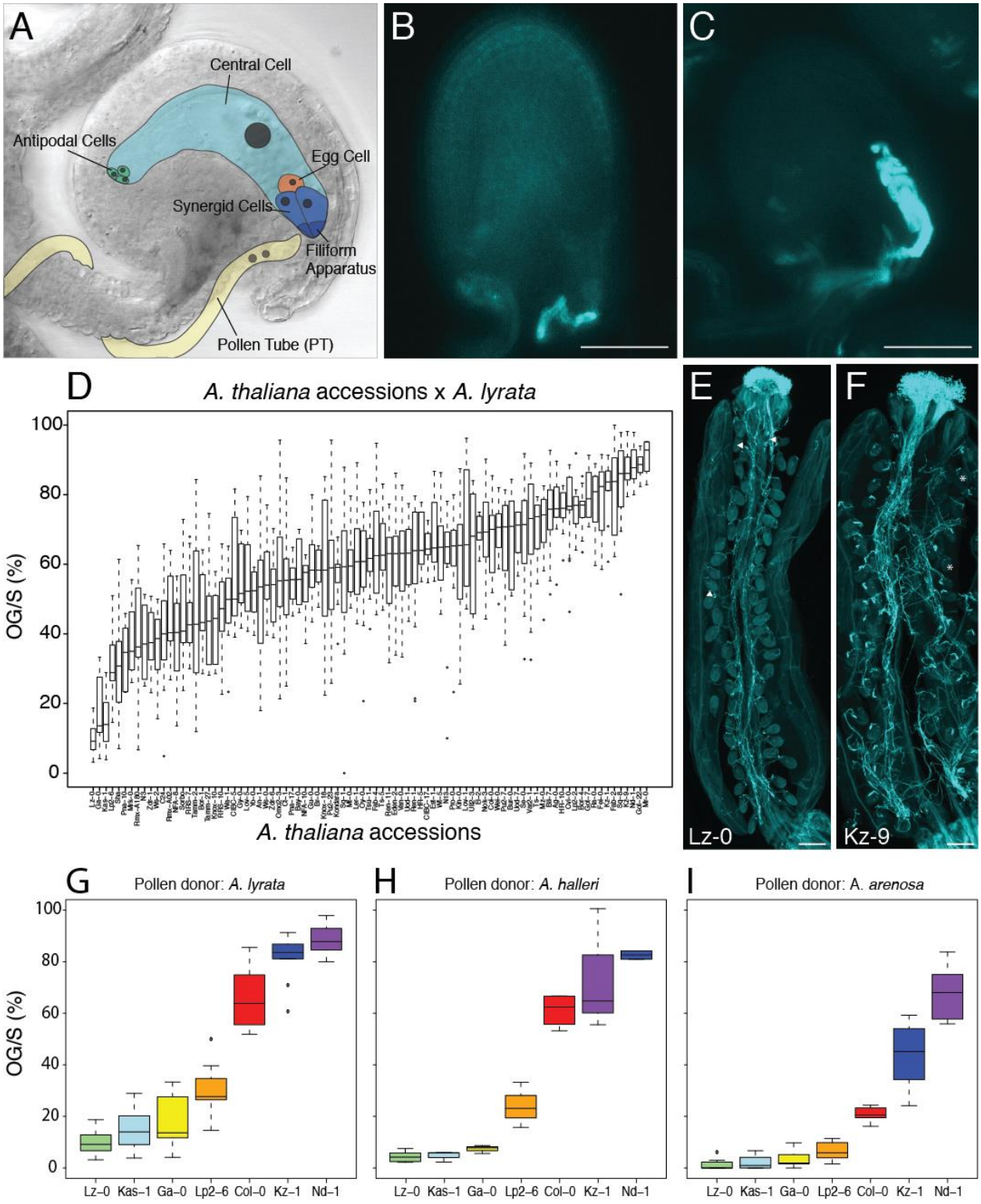


Figure 3.1: PT reception in interspecific crosses. (A) Diagram of the female gametophyte with its cell types. The synergids with their membrane-rich filiform apparatus are crucial for communication with the arriving PT. (B) Ovule with normal PT reception, visualized by callose staining of the PT cell walls with Aniline Blue. The PT stopped its growth and ruptured. (C) Ovule with PT overgrowth. The PT continues growing inside the female gametophyte. (D) Natural variation in the proportion of ovules with PT overgrowth per silique (OG/S) in 86 *A. thaliana* accessions that were pollinated with *A. lyrata* pollen. OG/S varies between 10% and more than 90%, depending on the genotype of the mother. **(continued next page)**

In interspecific crosses, *A. lyrata* PTs are normally guided to *A. thaliana* ovules, but upon arrival at the embryo sacs the PTs are not recognized and fail to arrest growth and discharge their sperm (*11*). This results in continuous growth of the PTs inside the unfertilized embryo sacs (Fig. 3.1C). Such PT overgrowth has been observed in a variety of interspecific pollinations between closely related species of the Brassicaceae and Ericaceae, respectively (*11, 12*); however, its molecular basis is not understood. Interspecific PT overgrowth phenocopies the female gametophytic mutants *feronia (fer), lorelei (lre), nortia (nta), turan (tun),* and *Zea mays embryo sac 4 (ZmES4)* RNAi-lines (*11*, *13-17*), which are defective in the reception of intraspecific PTs, caused by the disruption of communication between the gametophytes. In addition, *FER* has been proposed to be involved in interspecific PT recognition (*11*), and there is evidence that ZmES4 is sufficient to trigger PT growth arrest and rupture in a species-preferential manner (*17*).

To analyze interspecific hybridization barriers within the genus *Arabidopsis*, we assessed PT overgrowth in 86 *A. thaliana* accessions that were pollinated with *A. lyrata* pollen (table S1). PTs were visualized by staining callose in PT cell walls with Aniline Blue. We scored the proportion of ovules that failed to recognize interspecific PTs - leading to PT overgrowth - in relation to the total number of ovules that attracted a PT in a silique (overgrowth per silique, OG/S). We found a striking variation in the ability to recognize interspecific PTs between different *A. thaliana* accessions, with OG/S ranging from about 10-95% (Fig. 1D, broad-sense heritability H^2=0.7). Examples of accessions with extreme phenotypes are Lz-0 (10% OG/S) and Kz-9 (87.3% OG/S) (Fig. 3.1, E and F). There is no obvious correlation between the geographical origin of the accessions and their phenotype (fig. S2).

Figure Legend 3.1 (continued)
(**E**) A silique of Lz-0 pollinated with *A. lyrata* pollen. Most of the ovules show normal PT reception. Ovules with PT overgrowth are marked with an arrowhead. (**F**) A silique of Kz-9 pollinated with *A. lyrata* pollen. Most of the ovules display PT overgrowth. Asterisks mark ovules with normal PT reception. (**G**) A subset of *A. thaliana* accessions pollinated with *A. lyrata* pollen. (**H and I**) The same subset pollinated with pollen from *A. halleri* (H) and *A. arenosa* (I). The accessions show comparable OG/S with all three interspecific pollen donors. Scale bars: 50 μm (B and C), 250 μm (E and F).

To analyze whether the variation in the ability to recognize interspecific PTs is species-dependent, we pollinated a subset of *A. thaliana* accessions with low or high OG/S in crosses with *A. lyrata* (Lz-0, Kas-1, Ga-0, Lp2-6 and Col-0, Kz-1, Nd-1, respectively) also with pollen of *A. halleri* and *A. arenosa*. Whereas OG/S in the accessions pollinated with *A. lyrata* or *A. halleri* pollen was highly comparable (Fig. 3.1, G and H), the values were slightly lower for all accessions when pollinated by *A. arenosa* (Fig. 3.1I), indicating that *A. thaliana* recognizes *A. arenosa* PTs better than those of *A. lyrata* or *A. halleri*. However, accessions showing very low or high OG/S in crosses with *A. lyrata*, respectively, displayed a similar phenotype with *A. halleri* and *A. arenosa* as pollen donors, suggesting a common molecular PT reception mechanism for all three species. Thus, PT overgrowth is a hallmark of interspecific crosses with close Brassicaceae relatives and not a species-specific feature of *A. thaliana* and *A. lyrata*.

To investigate whether intraspecific PT reception was affected in accessions with high OG/S (Col-0, Kz-1, Kz-9, Nd-1, Fei-0 and Sq-8) we crossed them with *A. thaliana* pollen (from both low and high OG/S accessions). Intraspecific PT reception was normal in all tested accessions (fig. S3), indicating that high OG/S frequencies result from a failure in the recognition of interspecific PTs only, and are not due to a general defect in PT reception.

In order to identify loci causing the variation in interspecific PT reception in *A. thaliana*, we used publicly available SNP data from the 86 accessions to perform a genome-wide association study (GWAS) (*18*). To date, most GWAS in *Arabidopsis* have identified previously known candidate genes, with only a few studies identifying novel regulatory genes (*19, 20*). Applying the GLM function implemented in TASSEL (*21*), we identified a region on chromosome 1 containing 8 of the top 20 SNPs with the highest correlation to the OG/S trait (Fig. 3.2A and table S2). This 28kb region (positions 22814316 to 22842689) contains 6 genes and one pseudogene (Fig. 2B). Interestingly, calculation with mixed linear models that simultaneously correct for population structure and unequal genetic relatedness between individuals masked the peak (fig. S4).

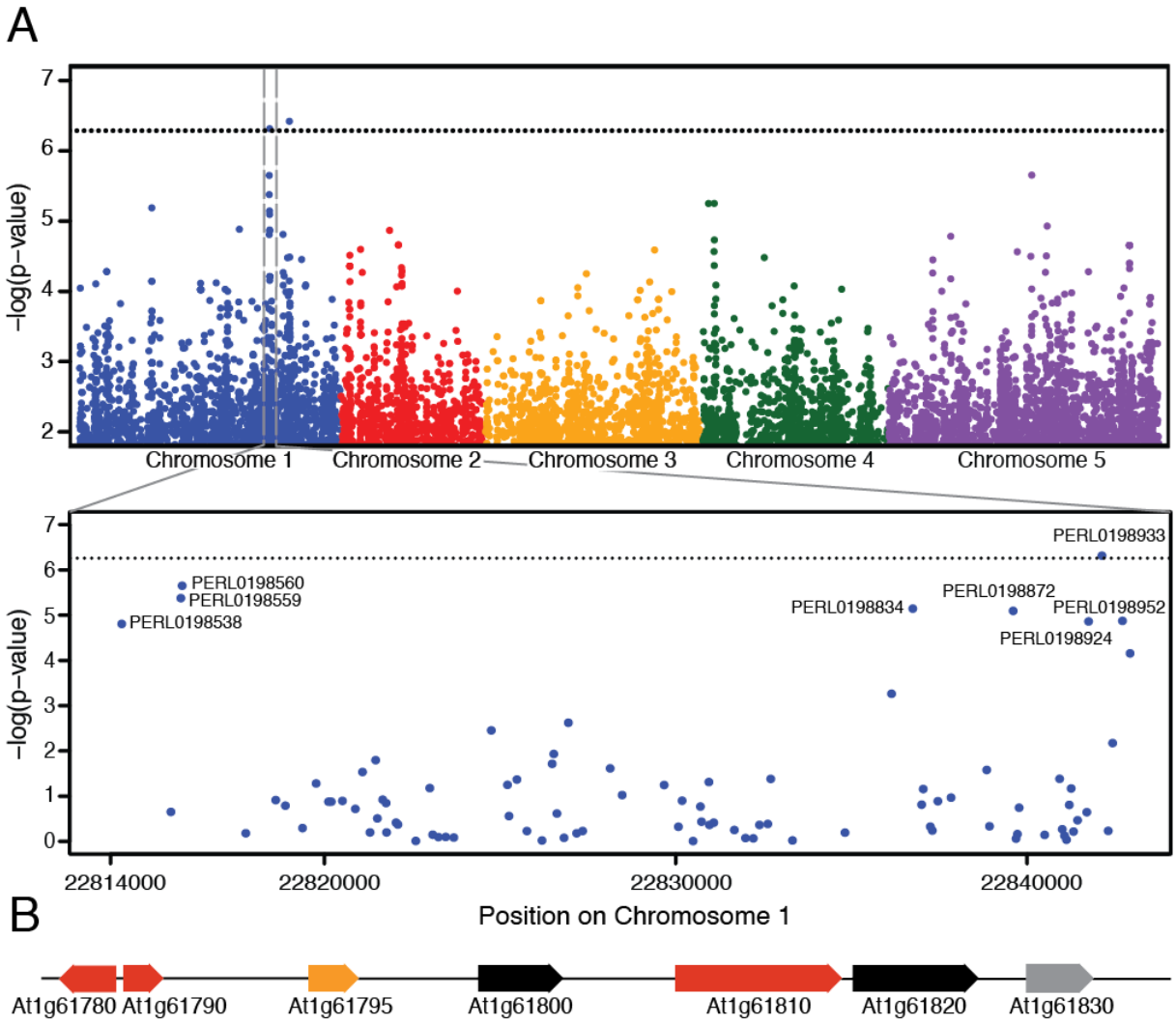
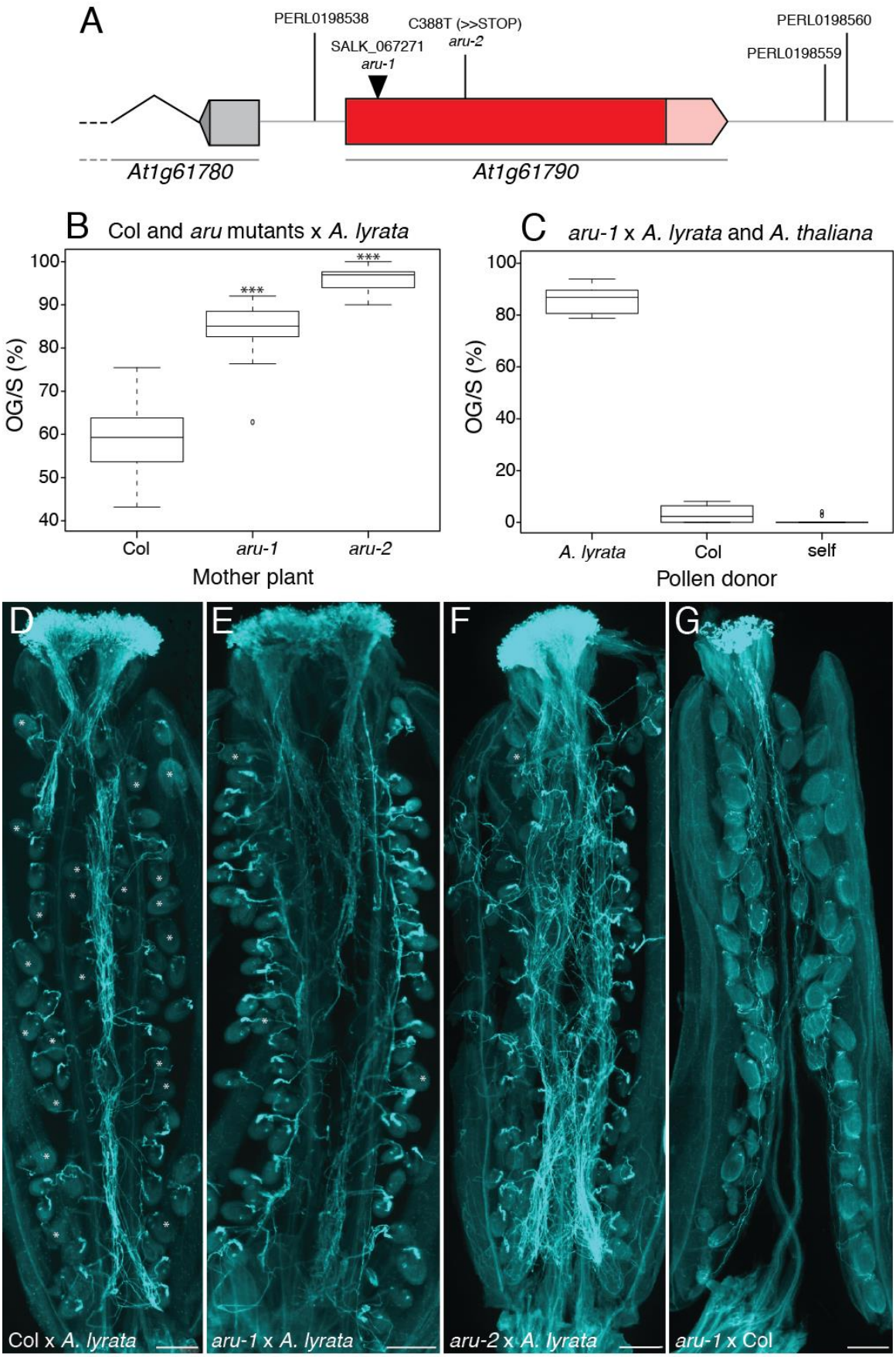


Figure 3.2: GWAS identifies an associated region on chromosome 1. (**A**) Manhattan plot showing a peak on chromosome 1 (grey box) with its highest correlated SNP showing significance at p<0,1 (after Bonferroni correction). The peak corresponds to a 28kb region spanning position 22,842,689-22,814,316 (magnified in the second panel). The 8 SNPs that were identified to be among the 20 most highly correlated ones in the GWAS are annotated with their PERL identifiers. (**B**) Genes and peudogenes (grey) in the 28kb region. Genes expressed in synergids are marked in red, genes without available expression data in orange (*22*).

To narrow down the 28kb candidate region to a single gene, we analyzed OG/S in T-DNA insertion lines of 4 synergid-expressed genes (*22*) in this region (*At1g61780, At1g61790, At1g61795, At1g61810*; fig. S5). In a homozygous T-DNA insertion allele disrupting the coding sequence of *At1g61790* (Fig. 3.3A), an average of 84.3% of ovules display *A. lyrata* PT overgrowth, significantly more than in the Col-0 wild type (58.7% OG/S, $p<0.01$, Fig. 3.3, B, D and E). We named the *At1g61790* gene *ARTUMES* (*ARU)* after the Etruscan goddess of night, nature and fertility (*23*), and the T-DNA allele was denoted *aru-1.* A second mutant allele, *aru-2*, carrying an EMS-induced premature stop-codon after aa residue 129 (*24*), also showed an increase in interspecific OG/S (96.1%, Fig. 3.3, B and F). In contrast, *aru* mutant ovules have no problems recognizing and receiving intraspecific PTs from *A. thaliana* (Fig. 3, C and G), indicating that the PT reception pathway is fully functional.

To confirm that *ARU* function is required for interspecific PT recognition in the synergids, we expressed an ARU-GFP translational fusion protein under the control of the *MYB98* and *FERONIA* promoters (*pMYB98::ARU-GFP* and *pFER::ARU-GFP*) in *aru-1* mutants. These promoters are highly active in synergids (*11*, *25*), and in ovules ARU-GFP was exclusively detected in these cells (Fig. 3.4A and fig. S6). ARU-GFP localized to perinuclear structures resembling the ER in synergids (Fig. 3.4A, inset), and co-localized with an ER-marker in transiently transformed onion epidermal cells (fig. S6). These results are consistent with the previously reported ER-localization of ARU-GFP in infiltrated tobacco leaves (*26*).

Figure 3.3: *artumes* mutants are impaired in interspecific, but not intraspecific PT reception. (**A**) Genomic region of *ARTUMES (At1g61790)* with the two mutant alleles *aru-1* and *aru-2*, and the surrounding polymorphisms identified by GWAS. (**B**) PT overgrowth of Col-0 wild-type and *aru* mutant plants in interspecific crosses. Both *aru* mutant alleles show significantly higher proportions of ovules with PT overgrowth per silique (OG/S, *** p<0,01). (**C**) PT overgrowth of *aru-1* in inter- and intraspecific crosses. The mutant is impaired in interspecific crosses with *A. lyrata* pollen, but not in intraspecific crosses with Col-0 or self pollen. (**D**) A silique of Col-0 pollinated with *A. lyrata* pollen. Ovules with normal PT reception (marked with asterisks) and with PT overgrowth are visible. (**E and F**) *aru-1* and *aru-2* siliques pollinated with *A. lyrata* pollen. Both mutant alleles show high proportions of ovules with PT overgrowth in interspecific crosses. Ovules with normal PT reception are marked with asterisks. (**G**) A silique of *aru-1* pollinated with intraspecific Col-0 pollen. All ovules display normal PT reception. Scale bars: 250 μm.


A
PERL0198538
C388T (>>STOP)
SALK_067271
aru-1
aru-2
PERL0198560
PERL0198559
At1g61780
At1g61790
B
Col and aru mutants x A. lyrata
OG/S (%)
100
90
80
70
60
50
40
***
***
Col
aru-1
aru-2
Mother plant
C
aru-1 x A. lyrata and A. thaliana
OG/S (%)
100
80
60
40
20
0
A. lyrata
Col
self
Pollen donor
D
E
F
G
Col x A. lyrata
aru-1 x A. lyrata
aru-2 x A. lyrata
aru-1 x Col

Mutant *aru-1* plants expressing a functional copy of ARU-GFP in their synergids displayed wild-type-like PT reception in interspecific crosses (Fig. 3.4B and fig. S6), indicating that *ARU* expression in synergid cells is sufficient to complement the *aru* mutant phenotype. Consistent with this, the ARU-GFP translational fusion protein driven by the endogenous promoter (*pARU::ARU-GFP*) is highly expressed in wild-type synergids (fig. S6), suggesting an important role for *ARU* in these cells.

We sequenced *ARU* in selected accessions with extreme OG/S in interspecific crosses (Lz-0, Ga-0, Kas-1, Lp2-6 and Col-0, Fei-0, Kz-1, Kz-9, Nd-1, Sq-8) to determine if differences in the protein sequence could explain the phenotypic variation. However, within this subset of accessions we could not find such differences. The protein sequences are very similar and, while amino acid exchanges in single accessions were found, no general pattern emerged (fig. S7). Alternatively, differential expression levels could cause the observed phenotypic variation. We used RNA extracts from pistils two days after emasculation from selected accessions (Lz-0, Kas-1, Ga-0 and Col-0, Fei-0, Kz-1, Kz-9) for quantitative RT-PCR using gene-specific primers. We found *ARU* mRNA levels to be similar between accessions (fig. S7) but because we used RNA from whole pistils, we cannot exclude the possibility that *ARU* is differentially expressed in synergids only, where it could directly influence interspecific PT reception.

Genes involved in reproductive isolation and speciation are often subject to selective pressures driving rapid divergence (*27*). We tested *ARU* plus 1000 bp up- and downstream sequence for signatures of positive selection by estimating Tajima's D (*28*) for a set of 96 *A. thaliana* accessions (*29*), including all accessions used in this study. A negative D is due to an excess of low frequency polymorphisms that can be caused by positive selection on the locus or by population expansion. Tajima's D was -2.07 for the 1000 bp upstream of the translation start, -1.57 for the coding sequence, and -1.58 for the 1000 bp downstream of the *ARU* stop-codon. All values significantly deviate from the neutral model ($p<0,05$) but do not fall into the 5%-tail of the estimated genomic distribution of D in *A. thaliana* (cut-off value: -2.08, (*29*)). Although the 1000 bp upstream region was very close to this cut-off, this indicates that the observed negative values might be influenced by demographic factors that shaped the entire genome rather than selective pressure acting on the *ARU* locus. Additionally, we estimated Fay and Wu's H, another test statistic to detect positive selection (*30*), which is not as sensitive to demographic factors as Tajima's D (*31*). All values for H were strongly

negative (upstream region: -20.20, coding sequence: -21.79, downstream region: -20.73, p<0,02), indicating that positive selection may indeed act on each of these regions. Thus, as in animal speciation genes, selective pressures likely contribute to shaping the genetic basis that underlies the phenotypic variation in interspecific PT reception in *A. thaliana*.

ARU encodes the OST3/6 subunit of the hetero-oligomeric plant oligosaccharyltransferase complex (OST), which catalyzes the co- or posttranslational transfer of pre-assembled carbohydrate oligomers ($Glc_3Man_9GlcNAc_2$) to asparagine (N) residues of polypeptides. N-glycosylation affects the substrate protein's folding, targeting, and/or processing through the ER. Subsequently, the N-linked glycan can be modified in the Golgi apparatus in a cell-type and species-specific manner, accounting for the functionality and binding specificity of the glycoprotein (*32*). The yeast OST consists of 8 subunits and the homologs of OST3/6, Ost3p and Ost6p, differ in their protein substrate and site-specific glycosylation efficiency (*33, 34*).

In plants, OST3/6 confers similar substrate specificity since in the *A. thaliana ost3/6 (aru)* mutant only a subset of glycoproteins is misglycosylated and therefore non-functional (*26*). Among these are the pathogen-associated molecular pattern (PAMP) receptor kinase EF-TU RECEPTOR (EFR), and KORRIGAN1, an endo-β-1,4-glucanase involved in cellulose biosynthesis (*26*). In line with this, *aru* has previously been identified in an EMS-screen for cell wall mutants (*24*). Some of the known members in the PT reception pathway, *FER* and *NTA*, have been implicated in the perception of cell wall perturbations, pathogen resistance, and innate immunity (*15, 35, 36*). Moreover, FER (a receptor-like kinase with an extracellular malectin-binding domain) and LRE (a glycosylphosphatidylinositol-anchored protein) are likely to be glycosylated (*11, 14, 37*) and could be substrates of ARU. To test this, we analyzed the expression and localization of fluorescent FER and LRE fusion proteins in *aru* ovules. We included NTA, which itself does not contain any putative glycosylation site but whose subcellular localization depends on *FER* signaling (*15*). All fusion proteins displayed a wild-type-like subcellular localization in the synergids of *aru* embryo sacs (fig. S8). In addition, a band-shift assay to assess the glycosylation state of FER-GFP did not show any size differences in *aru* compared to wild-type inflorescences (fig. S9), indicating that FER is not differentially glycosylated in *aru*. However, we cannot rule out that only few specific FER

glycosylation sites remain unglycosylated in *aru,* which may allow the protein to recognize intra- but not interspecific PTs.

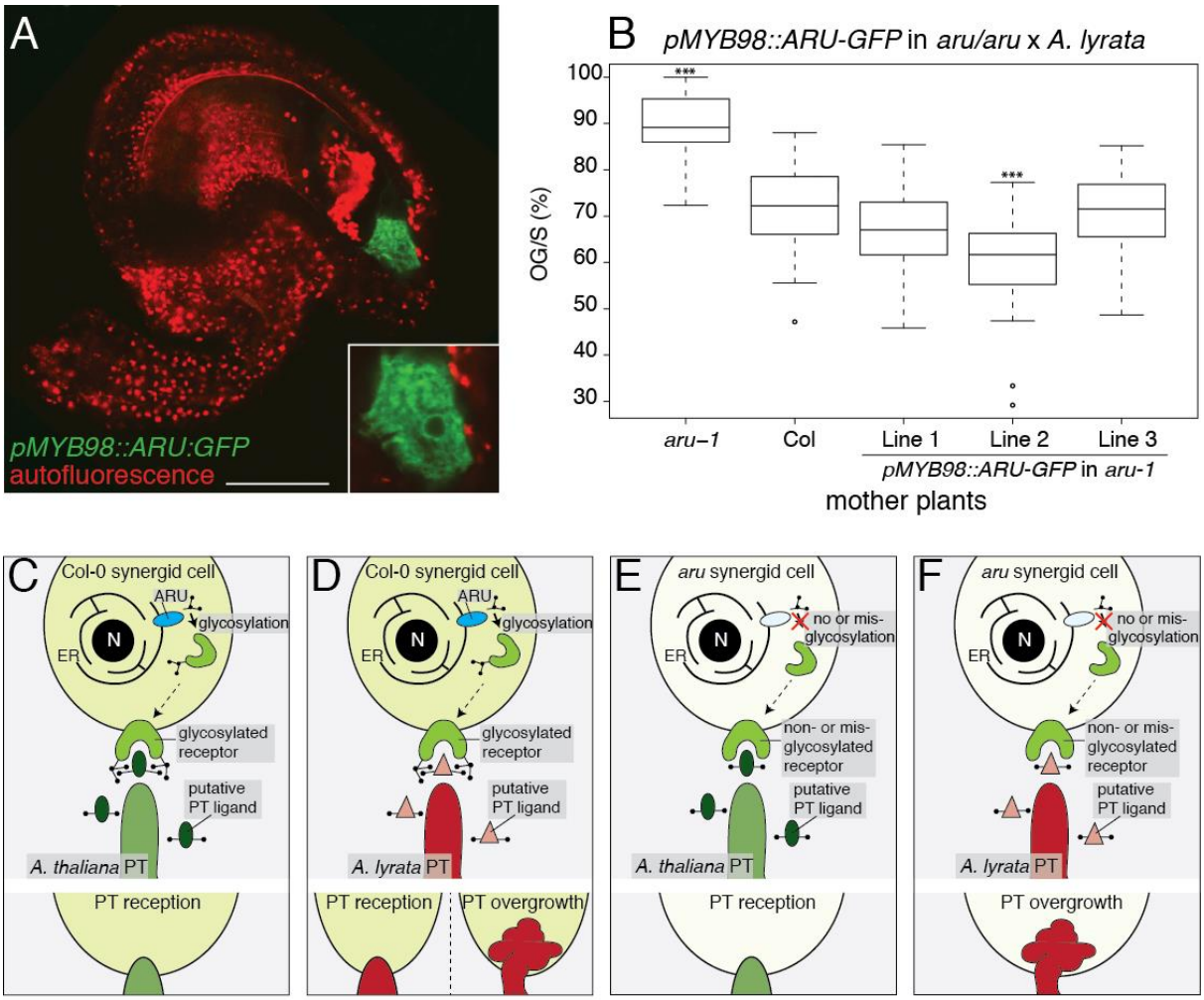


Figure 3.4: Synergid-specific expression of *ARTUMES* can complement the mutant phenotype. (A) An ovule expressing *pMYB98::ARU-GFP* in the synergids. Inset: ARU-GFP localizes to perinuclear structures resembling the ER. (B) PT overgrowth in interspecific crosses using *aru-1*, Col-0, and three independent transformant lines of *pMYB98::ARU-GFP* in the *aru-1* background as mother plants and *A. lyrata* as pollen donor. All three transformant lines complement the mutant phenotype; line 2 shows even lower OG/S than the wild type. Significance levels in comparison to Col-0 (*** p<0,01). Scale bar: 50 µm.

A possible interpretation of our results is that FER, or a so far unknown (co-) receptor, binds putative ligands from intraspecific PTs both via specific interactions with carbohydrates on the receptor protein and via direct protein-protein interactions, a mechanism similar to the proposed "domain-specific model" in mammalian sperm-egg binding (Fig. 3.4C, (*38*)). Ligands from interspecific PTs might not be able to sufficiently interact via protein-protein interactions alone and thus could be only partially

recognized via the carbohydrate moieties, explaining the partial *A. lyrata* PT reception success in Col-0 ovules (Fig. 3.4D). In *aru*, and potentially also in *A. thaliana* accessions with a similar phenotype, changes in the glycosylation status of the receptor could completely abolish the ability to recognize and receive interspecific PTs (Fig. 3.4F) while ligands from *A. thaliana* PTs are still efficiently recognized via protein-protein interactions, leading to normal PT reception (Fig. 3.4E).

The crosstalk between gametophytes constitutes a specific form of cell-cell communication. Cellular interactions are often mediated by specific binding of an extracellular ligand to a receptor, triggering downstream signaling cascades in the recipient cell. Most extracellular ligands and receptors are heavily glycosylated (*39*), which influences their binding specificities and conformation, such that already the absence of a single glycosylation motif can reduce or abolish a receptor's function and ligand-binding affinity (*40, 41*). Our results suggest that both protein-protein interactions and recognition mediated by carbohydrates are crucial factors to ensure species-specific PT reception. Thus, divergent evolution of receptor-ligand pairs as well as of the factors controlling their glycosylation status could establish new species barriers. Deciphering the molecular basis of speciation in plants might enable us to overcome existing hybridization barriers, which could eventually be of great agronomic importance.

Materials and Methods

Plant material and growth conditions

The *A. thaliana* accessions were part of the Nordborg collection for GWAS (*18*). Amplified seed stocks were kindly donated by Ortrun Mittelsten Scheid (Gregor Mendel Institute, Vienna). After stratification (two days at 4°C), the seeds were allowed to germinate for 6 days on MS plates (22°C, 16h light, MS from Carolina Biological Supply). Because some accessions require vernalization, all seedlings were kept in a vernalization chamber for 5 weeks (4°C, 16h light) on MS plates before they were transferred to soil (ED73, Universalerde).

The accessions were grouped into early- (4 incomplete blocks A, B, C and D), mid-, and late flowering plants (3 complete blocks A, B, C each) according to the flowering time

assessed in (*18*), and were grown in a greenhouse chamber (22°C, 16h light) in an incomplete randomized block design. See table S1 for the assignment of accessions to the blocks.

A. lyrata (*11*), *A. halleri* (a gift from Marcus Koch, University of Heidelberg) and *A. arenosa* (donated by Matthias Helling, University of Zurich) plants were stratified for 10 days and grown in the same greenhouse chamber. Plants were vernalized to induce flowering (see above).

SALK-lines were obtained from NASC. The EMS allele *aru-2 (24)* was a gift from Peter McCourt (University of Toronto). The plants were grown as described before (*15*). See table S3 for a complete list of mutants analyzed in this study.

Crosses and Aniline Blue staining

Flowers were emasculated and the pistils were pollinated two days after emasculation. Siliques were collected two days after pollination and fixed for Aniline Blue staining in 9:1 Ethanol:Acidic Acid. Aniline Blue staining was performed as described previously (*15*), and the samples were analyzed at the Leica DM6000B microscope (Leica Microsystems). For GWAS phenotyping, 9-20 siliques of a minimum of 3 individuals were analyzed for each accession (Exceptions: Zdr-1: 7 siliques, Got-7: 4 siliques from 2 individuals).

GWAS

Association mapping was conducted using the mean values of the proportions of ovules with PT overgrowth per silique (OG/S) as phenotypes. An *A. thaliana* 250K Affymetrix SNP genotyping dataset (*18*) was downloaded from https://cynin.gmi.oeaw.ac.at/home/resources/atpolydb. GWAS analyses were performed using a compressed mixed linear model, using population parameters previously determined (*45, 46*) and a kinship matrix to account for family relatedness, in the R package GAPIT (*42*). The mixed linear models were run with and without principal components as fixed effects to correct for population structure. Multiple testing was controlled using the Bonferroni correction and false-discovery rate (*47*). GWAS analyses were also run using a general linear model in the web-based interface TASSEL3.0 (*21*) and an accelerated mixed model with Box-Cox transformed phenotypes in GWAPP (*43*).

Constructs for stable plant transformation

For *pMYB98/pFER::ARU-GFP* the complete coding sequence of *ARU* without the stop-codon was amplified using gene-specific primers with attB-sites for Gateway cloning: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGCTCAAATCAAAACTCGTC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCACGCCAACTCGATGGCCAATACGGA-3'. We introduced the PCR-fragment into pDONR207, and subsequently into the destination vector (using the *E. coli* strain DH5-alpha F'I^q from New England Biolabs). The destination vector was a modification of the plant Gateway vector pMDC83, which contains the 2x35S-promotor before and *GFP* after the Gateway cassette (*48*). For our purpose, we exchanged the original 2x35S-promoter with the promoters of *FER* (*11*) and *MYB98 (25)* to express *ARU* specifically in synergid cells. The *MYB98* promoter was amplified from Col-0 with primers 5'-TTTAAGCTTATACACTCATTGTCCTTCG-3' and 5'-CCCTCTAGATGTTTTGGAAAGGAGAAAAAA-3', introducing a *Hin*dIII and *Xba*I restriction site, respectively. The *FER* promoter was amplified from the *pFER::FER-GFP* construct (*11*) using specific primers 5'-TTTGGTAAGCTTCGATTTAAGCGAG-3' and 5'-TTTTCTAGACGATCAAGAGCACTTCTCCGGG-3', which introduce *Hin*dIII and *Xba*I restriction sites as well. The 2x35S-promoter was cut out of pMDC83 (*48*) with *Hin*dIII/*Xba*I (New England Biolabs), and the PCR fragments were introduced into the dephosphorylated vector backbone by ligation. pDONR207 carrying the *ARU* coding sequence and the modified destination vector were combined in an LR reaction. The resulting vectors, *pFER::ARU-GFP* and *pMYB98::ARU-GFP* were transformed into *Agrobacterium tumefaciens* strain CV1310, and homozygous *aru-1* plants were transformed by the floral dip method (*49*). The complementation assays were conducted in the T2 and T3 generations with plants homozygous for the complementation construct and the *aru-1* mutation.

pARU::ARU-GFP was cloned by amplification of a part of *ARU* coupled to *GFP-tNOS* from *pFER::ARU-GFP* with primers 5'-GCGTTAACGCTTTACCTCA-3' (including the natural *Hpa*I site in *ARU*) and 5'-TTTGGATCCAGTAACATAGATGACACCGCG-3' (introducing a *BamH*I site after tNOS). This fragment was introduced by ligation into the pMDC99 vector (*48*), carrying the genomic fragment of *ARU* including 1492 bp of upstream and 865 bp of downstream sequence (amplified from Col-0 genomic DNA with primers 5'-TTTTACTAGTAGGCAATTCCATCAGTTGTT-3' and 5'-TTTTGGTACCGTTACTTCACTTTCTCGAGT-3', introducing *Spe*I and *Kpn*I sites,

respectively). By cutting this vector with *Hpa*I and *BamH*I, part of the *ARU* coding sequence and the downstream sequence were removed and replaced with the respective fragment of *ARU* coupled to *GFP-tNOS*, resulting in *pARU*(1492bp)*::ARU-GFP*, which was transformed into Col-0.

pLRE::LRE-Citrine was cloned with overlapping PCR fragments that were assembled using the Gibson cloning Master Mix from New England Biolabs according to the manufacturer`s recommendations. The 779bp long promoter sequence with the predicted signal peptide from *LRE* (*14*) was amplified with primers 5'-GTGCTGCAAGGCGATTAAGTCCGTGTGCTCTGTCTGCATT-3' and 5'-CACAGCTCCACCTCCACCTCCAGGCCGGCCTATGGAACTTGAAGAGGAGAGAGA-3', introducing an overhang complementary to the vector pMDC99 (*48*). *Citrine* was amplified from the transgenic line CS36962 (ordered from Arabidopsis Biological Recource Center, ABRC), using gene-specific primers with overhang primers for the signal peptide of *LRE* and overhang primers for the GPI-anchor of *LRE*: 5`-GGCCGGCCTGGAGGTGGAGGTGGAGCTGTGAGCAAGGGCGAGGAGCT-3' and 5'-GGCCCCAGCGGCCGCAGCAGCACCAGCAGGATCCTTGTACAGCTCGTCCA-3'. The GPI-anchor of *LRE* was amplified with overhang primers for pMDC99: 5'-TGCTGGTGCTGCTGCGGCCGCTGGGGCCTCGGGTATGTCTTTTTGTTGTC-3' and 5'-AGCTCCACCGCGGTGGCGGCCGCTCTAGAAGTCTCGCTTCTTCTTTTGT-3'. pMDC99 was amplified with overhang primers for the *LRE* promoter and the GPI-anchor using primers: 5'-ACTTAATCGCCTTGCAGCAC-3' and 5'- TCTAGAGCGGCCGCCACCGCGG-3'. All constructs were verified by sequencing. *pFER::FER-GFP* and *pNTA::NTA-GFP* were described previously (*11, 15*).

ARU-GFP subcellular localization

We used the *pFER::ARU-GFP* construct for microprojectile bombardment of onion epidermal cells and co-localized it with the ER-marker pER-rk (mCherry) obtained from the ABRC (*44*). Biolistic bombardment of onion epidermis was performed as described (*15*).

For visualizing GFP expression in the synergids, flowers were emasculated and pistils were dissected two days after emasculation (dae) to ensure the development of mature,

unfertilized embryo sacs. The tissue was mounted on slides in 1M Glycine, pH 9,6. Images were captured on Leica Confocal Microscopes SP2 or SP5 (Leica Microsystems).

RNA extraction and RT-PCR

RNA from pistils (25 pistils, 2 dae) and inflorescences was extracted using the Trizol reagent (Invitrogen) according to the manufacturer's recommendations. cDNA was reverse transcribed using Oligo-dT primers and Superscript II reverse transcriptase from Invitrogen.

RT-PCR of *ARU* was done using primer 5'-CAATGTGCTTGTTCGAGTG-3' and 5'-ATCCAGTCTTCCAGTTATCCA-3'. For qRT-PCR of *ARU* in *A. thaliana* accessions, the primers 5'-GTTTGTTACCAATGTGCTTGTTCG-3' and 5'-TCCATATCCAGTCTTCCAGTTATCC-3' were used and expression levels were normalized against *UBIQUITIN C* (*UBC*, primers: 5'- ATGCTTGGAGTCCTGCTTGG-3' and 5'-TGCCATTGAATTGAACCCTCTC-3').

Population genetic analyses and statistical tests

Sequences of *ARU* and 1000 bp up- and downstream flanking regions were downloaded from http://signal.salk.edu/atg1001/3.0/gebrowser.php. For accessions for which no sequences or only sequences with missing data were available, we amplified the whole region from genomic DNA using Primers 5'-TTTGCTATAGGCACATGTGT-3' and 5'-GACCCGAAATTGTCAAATGA-3', and sequenced the resulting PCR products of Bay-0, Fab-2, Fab-4, Omo-2-3, Knox-10, Kz-1, LL-0, Lz-0, Mr-0, Mrk-0, Zdr-6. The upstream region was sequenced additionally from Got-7, Pu2-23 and Spr1-6 (primer 5'-TTTGCTATAGGCACATGTGT-3' and 5'-CGGAGGTTAGGAATTTTGAGA-3') and the downstream region from Got-7, Pu2-23, Kz-9, Mz-0, Pro-0, Van-0 and Var2-1 (primer 5'-CAATGTGCTTGTTCGAGTG-3' and 5'-GACCCGAAATTGTCAAATGA-3'). Tajima's D and Fay and Wu's H were calculated separately for the 1000 bp up- and downstream as well as the coding sequence with the set of 96 accessions (*29*) using DnaSP 5.10 (*50*). Several accessions that had big indels in the up- and downstream regions were left out from the analysis (Mr-0, Got-7, Pu2-23 and Spr1-6 for the upstream, Var2-1, Nok-3 and Got-7 for the downstream region). *A. lyrata* was used as outgroup. P-values against the null model

were obtained by running 10,000 coalescent simulations and for Tajima's D, the 5% quantile was calculated using the estimates for D given in (*29*).

Western blot and deglycosylation assay

Proteins were extracted from inflorescences by grinding them in a mixer mill and subsequently adding extraction buffer (50mM Tris pH7.5, 10mM NaCl, 0.5% Triton X-100 and a tablet of Complete Mini Protease Inhibitor Cocktail (Roche)). Extracts were incubated on ice for 15 min and centrifuged for 3 min at 14,000 rpm. Protein extracts were boiled at 95^{o}C with SDS loading buffer, and loaded on a 10% gel followed by SDS-PAGE under reducing conditions. Membranes were probed with the GFP B-2 antibody (Santa Cruz Biotech). EndoH (New England Biolabs) digestion was performed according to the manufacturer's instructions under reducing conditions.

References

1. A. Widmer, C. Lexer, S. Cozzolino, Evolution of reproductive isolation in plants. *Heredity* **102**, 31 (2009).
2. L. H. Rieseberg, B. K. Blackman, Speciation genes in plants. *Annals of Botany* **106**, 439 (2010).
3. G.M. Zinkl, B.I. Zwiebel, D.G. Grier, D. Preuss, Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine. *Development* **126**, 5431 (1999).
4. S. Okuda, H. *et al.,* Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**, 357 (2009).
5. M. L. Marton, A. Fastner, S. Uebler, T. Dresselhaus, Overcoming hybridization barriers by the secretion of the maize pollen tube attractant ZmEA1 from *Arabidopsis* ovules. *Curr Biol* **22**, 10 (2012).
6. A. Lausser, I. Kliwer, K. O. Srilunchang, T. Dresselhaus, Sporophytic control of pollen tube growth and guidance in maize. *J Exp Bot* **61**, 673 (2010).
7. M. M. Kanaoka *et al.,* Identification and characterization of TcCRP1, a pollen tube attractant from *Torenia concolor*. *Annals of Botany* **108**, 739 (2011).
8. T. Higashiyama *et al.*, Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*. *Plant Physiol* **142**, 481 (2006).
9. D. Lewis, L. K. Crowe, Unilateral interspecific incompatibility in flowering plants. *Heredity* **12**, 233 (1958).
10. S. A. Kessler, U. Grossniklaus, She's the boss: signaling in pollen tube reception. *Curr Opin Plant Biol* **14**, 622 (2011).
11. J. M. Escobar-Restrepo *et al.,* The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* **317**, 656 (2007).
12. V. R. Kaul, J. L. Rouse, E.G. Williams, Early events in the embryo sac after intraspecific and interspecific pollinations in *Rhododendron kawakamii* and *R. retusum*. *Canadian Journal of Botany* **64**, 282 (1986).

13. N. Huck, J. M. Moore, M. Federer, U. Grossniklaus, The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception. *Development* **130**, 2149 (2003).

14. A. Capron *et al.,* Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene. *Plant Cell* **20**, 3038 (2008).

15. S. A. Kessler *et al.,* Conserved molecular components for pollen tube reception and fungal invasion. *Science* **330**, 968 (2010).

16. H. Lindner *et al.,* SNP-Ratio Mapping (SRM): Identifying lethal alleles and mutations in complex genetic backgrounds by next-generation sequencing. *Genetics* **191**, 4 (2012).

17. S. Amien *et al.,* Defensin-like *Zm*ES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS Biol* **8** (2010).

18. S. Atwell *et al.,* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627 (2010).

19. P. E. Verslues, J. R. Lasky, T. E. Juenger, T. W. Liu, M. N. Kumar, Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis thaliana*. *Plant Physiol* **164**, 1 (2013).

20. M. Meijon, S. B. Satbhai, T. Tsuchimatsu, W. Busch, Genome-wide association study using cellular traits identifies a new regulator of root development in *Arabidopsis*. *Nat Genetics* **46**, 1 (2013).

21. P. J. Bradbury *et al.,* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633 (2007).

22. S. E. Wuest *et al., Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol* **20**, 506 (2010).

23. E. Gerhard, Über die Gottheiten der Etrusker. Druckerei der Königlichen Akademie der Wissenschaften, Berlin (1847).

24. R. S. Austin *et al.,* Next-generation mapping of *Arabidopsis* genes. *Plant J* **67**, 715 (2011).

25. R. D. Kasahara, M. F. Portereiko, L. Sandaklie-Nikolova, D. S. Rabiger, G. N. Drews, *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell* **17**, 2981 (2005).

26. A. Farid *et al.,* Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses. *Plant Physiol* **162**, 24 (2013).

27. W. J. Swanson, V. D. Vacquier, The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137 (2002).

28. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585 (1989).

29. M. Nordborg *et al.,*The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**, 7(2005).

30. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405 (2000).

31. K. J. Schmid, S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar, T. Mitchell-Olds, A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169**, 1601 (2005).

32. M. Aebi, N-linked protein glycosylation in the ER. *Biochimica et Biophysica acta* **1833**, 2430 (2013).

33. M. Schwarz, R. Knauer, L. Lehle, Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit. *FEBS letters* **579**, 6564 (2005).

34. B. L. Schulz, M. Aebi, Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-glycosylation efficiency. *Mol Cell Proteomics* **8**, 357 (2009).

35. N. F. Keinath *et al.,* PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. *J Biol Chem* **285**, 39140 (2010).

36. A. Y. Cheung, H. M. Wu, *THESEUS 1*, *FERONIA* and relatives: a family of cell wall-sensing receptor kinases? *Curr Opin Plant Biol* **14**, 632 (2011).

37. H. Lindner, L. M. Müller, A. Boisson-Dernier, U. Grossniklaus, *Cr*RLK1L receptor-like kinases: not just another brick in the wall. *Curr Opin Plant Biol* **15**, 659 (2012).

38. G. F. Clark, Molecular models for mouse sperm-oocyte binding. *Glycobiology* **21**, 3 (2011).

39. D. F. Zielinska, F. Gnad, K. Schropp, J. R. Wisniewski, M. Mann, Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol Cell* **46**, 542 (2012).

40. R. A. van der Hoorn *et al.,* Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *Plant Cell* **17**, 1000 (2005).

41. H. Haweker *et al.,* Pattern recognition receptors require N-glycosylation to mediate plant immunity. *J Biol Chem* **285**, 4629 (2010).

42. A. E. Lipka *et al.,* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397 (2012).

43. Ü. Seren *et al.,* GWAPP: a web application for genome-wide association mapping in *Arabidopsis*. *Plant Cell* **24**, 4793 (2012).

44. B. K. Nelson, X. Cai, A. Nebenfuhr, A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J* **51**, 1126 (2007).

45. Z. Zhang *et al.,* Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42**, 355 (2010).

46. H. M. Kang *et al.,* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348 (2010).

47. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289 (1995).

48. M. D. Curtis, U. Grossniklaus, A Gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiol* **133**, 462 (2003).

49. S. J. Clough, A. F. Bent, Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735 (1998).

50. J. Rozas, R. Rozas, DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Computer Applications in the Biosciences* **11**, 621 (1995).

Supplementary Figures and Tables

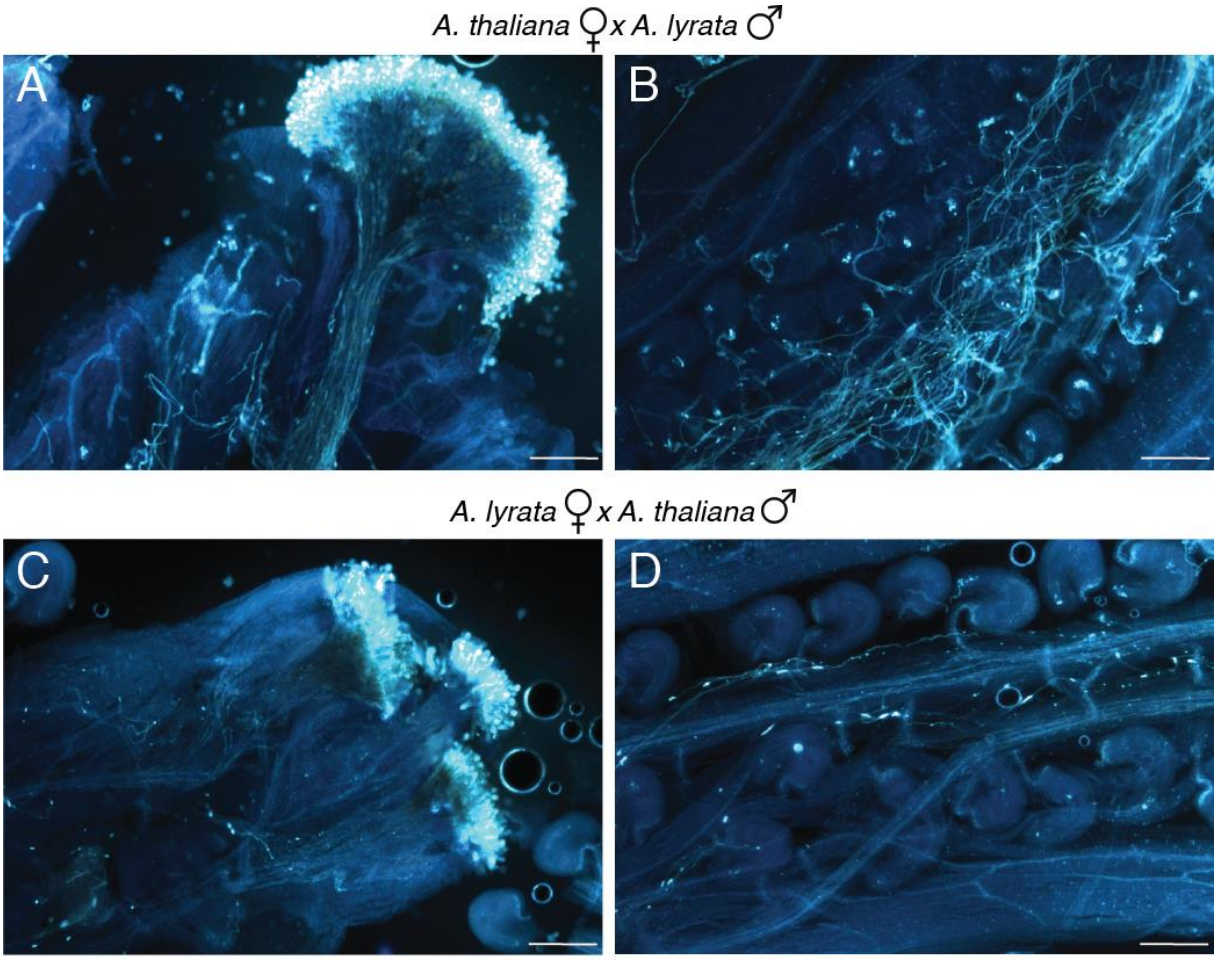


Figure S1: The nature of interspecific crossing barriers between *A. thaliana* and *A. lyrata* depends on the direction of the cross. (A and B) When *A. thaliana* pistils are pollinated with *A. lyrata* pollen, the PTs germinate (A) and grow through the transmitting tract towards the ovules (B). If they are not recognized correctly, the PTs continue growing inside the female gametophyte (PT overgrowth). (C and D) Pistils of *A. lyrata* pollinated with *A. thaliana* pollen. In this case, the crossing barrier occurs earlier than in the reciprocal cross: most pollen grains do not germinate on the stigma (C), and only a few are visible in the transmitting tract (D). Scale bar: 250 µm.

Figure S2: Map of the geographical origin of the *A. thaliana* accessions used for phenotyping with their color-coded phenotypes (% of ovules with PT overgrowth per silique). Map created with www.gpsvisualizer.com.

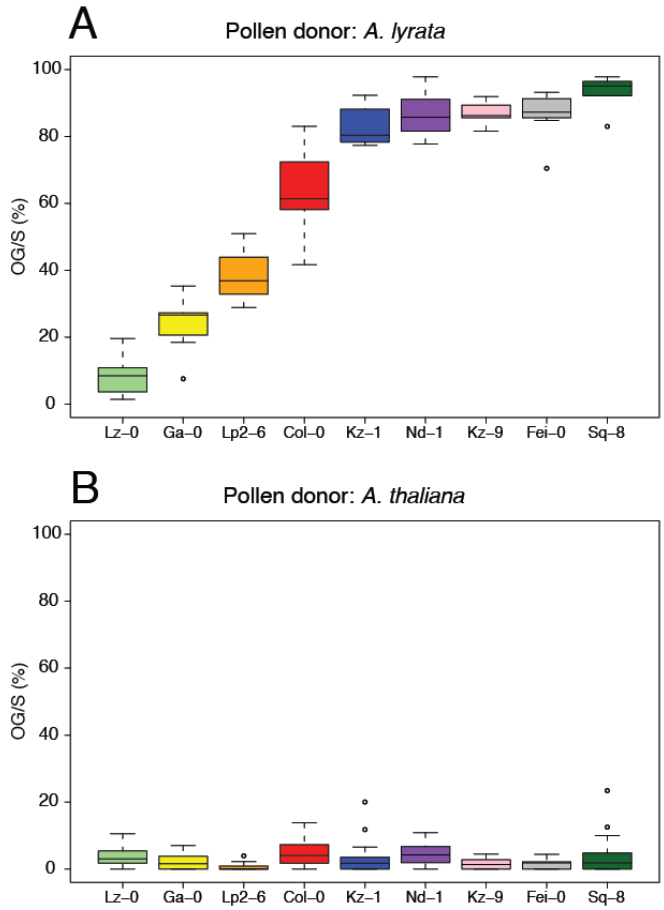


Figure S3: Intraspecific PT reception is normal in accessions with high OG/S in interspecific crosses. (A) PT overgrowth in a subset of accessions pollinated with *A. lyrata* pollen. (B) The same subset pollinated with intraspecific *A. thaliana* pollen. As no difference was observed between *A. thaliana* pollen from low or high OG/S accessions, the data were pooled. All accessions, no matter if they display low or high PT overgrowth phenotypes in interspecific crosses, show normal intraspecific PT reception with only a low level of PT overgrowth.

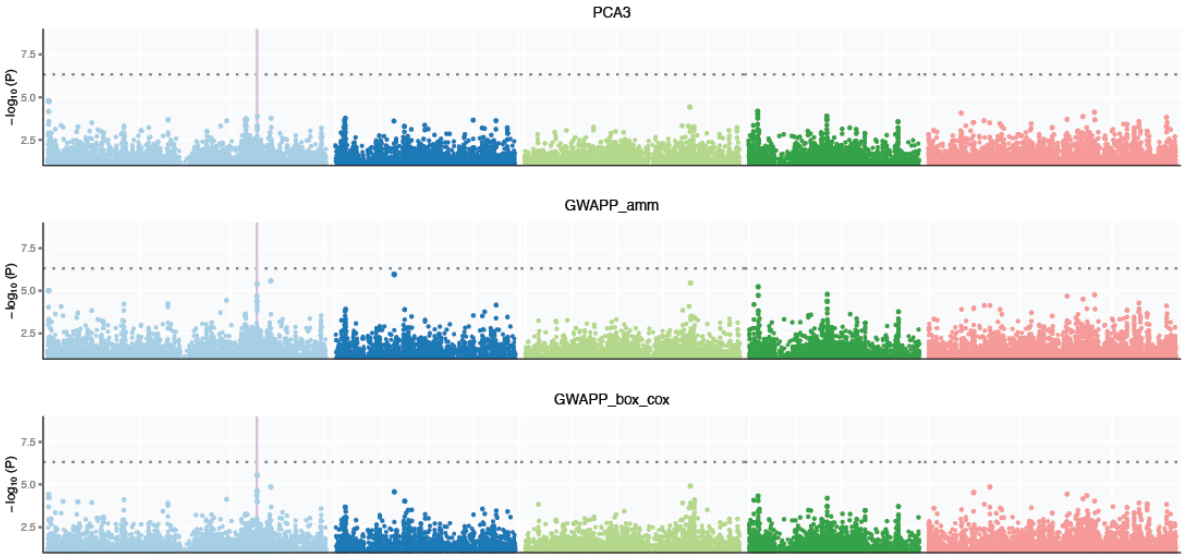


Figure S4: Manhattan plots of GWAS analyses with mixed models. First panel: GWAS calculated with GAPIT (PCA3, (*42*)), correcting for population structure and relatedness using principal component analysis (PCA). There is no peak at the *ARU* locus (vertical line). The second and third panels show outputs of the GWAS calculated with the web-program GWAPP (*43*). With both methods (accelerated mixed model, amm, and especially after box_cox normalization), a peak at the *ARU* locus is visible, although it is not significant. Dotted line: p=0,1 (Bonferroni corrected).

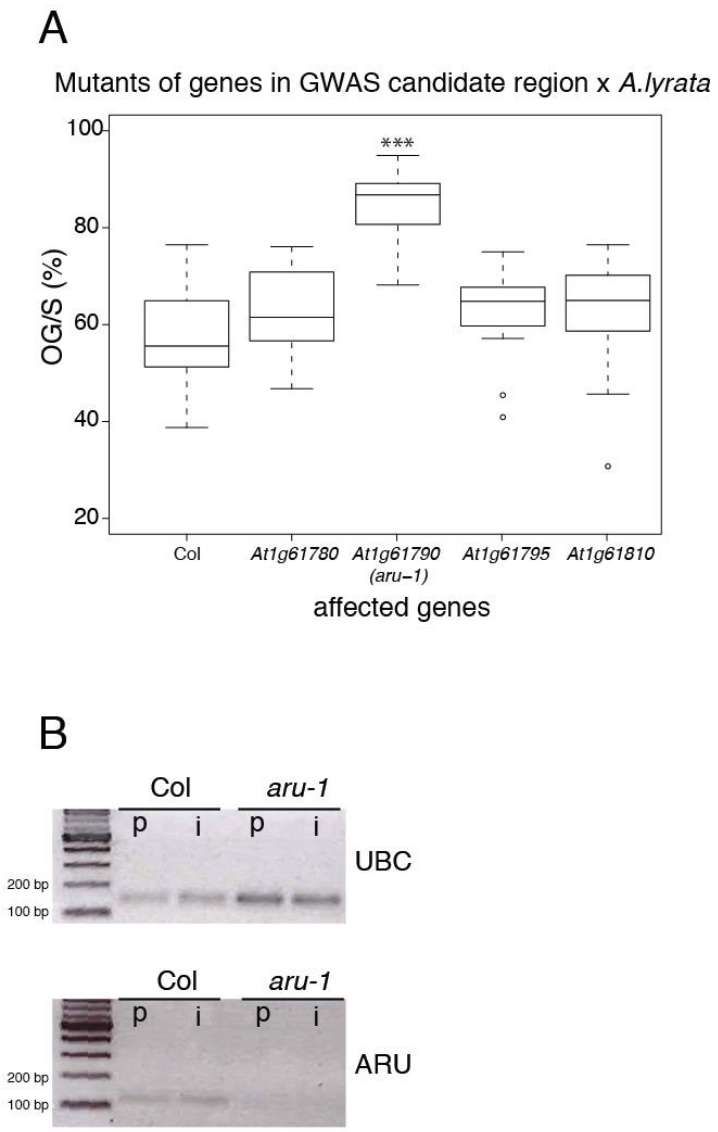


Figure S5: Mutant analysis of synergid-expressed genes in the GWAS-candidate region on chromosome 1. (A) PT overgrowth of candidate mutants in interspecific crosses with *A. lyrata*. Only *aru-1*, a T-DNA insertion in *At1g61790*, shows a phenotype different from wild-type Col-0 (*** p<0,01). (B) Semi-quantitative RT-PCR showing that *ARU* mRNA is absent in *aru-1*. The housekeeping gene *UBC* was used as control.

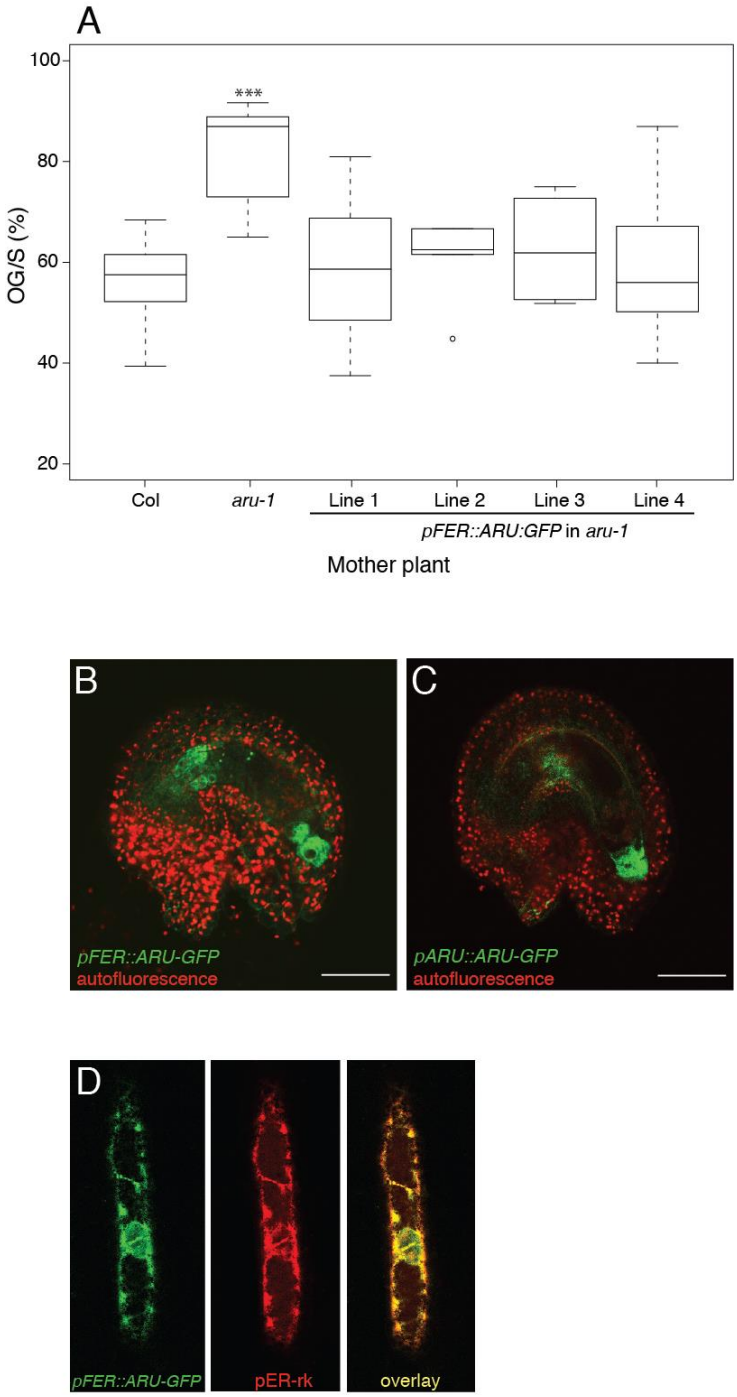


Figure S6: *ARU-GFP* expression under the control of the *FERONIA* promoter (*pFER*). (A) PT overgrowth in interspecific crosses with *A. lyrata*. Four lines expressing *ARU-GFP* under the control of the *FER* promoter in the *aru* background that complement the mutant phenotype. Significance levels in comparison to Col-0 (*** p<0,01). (B) An ovule expressing *pFER::ARU-GFP*. Expression is strongest in the synergids and ARU-GFP localizes to a perinuclear structure. (C) An ovule expressing *pARU::ARU-GFP* with strong expression in the synergids. (D) Transient expression of FER-GFP (green, from *pFER::ARU-GFP*) and the ER-marker pER-rk (red, (*44*)) in transiently transformed onion epidermal cells. Both markers co-localize, indicating that ARU localizes to the ER. Scale bar: 50 µm.

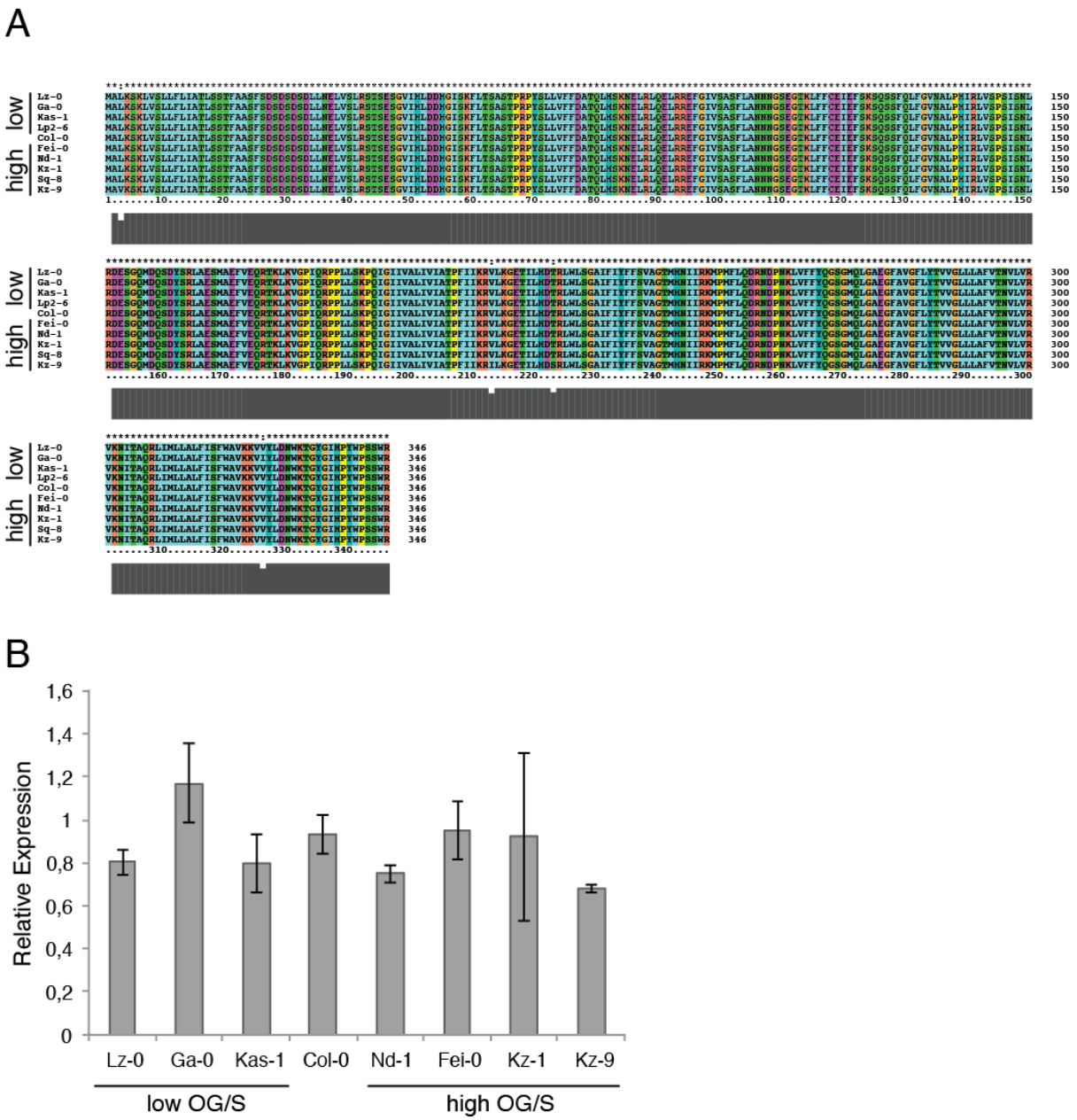


Figure S7: Alignment of ARU protein sequences and qRT-PCR of accessions with low and high PT overgrowth in interspecific crosses. (A) An alignment of ARU protein sequences of accessions with extreme OG/S phenotypes. No obvious correlation to the variation in the phenotype could be detected at the protein level. Sequences were obtained from the 1001 genomes project and our own sequencing with gene-specific primers, respectively. (B) Relative *ARU* mRNA levels in pistils of accessions with low and high proportions of PT overgrowth in interspecific crosses. All accessions showed similar expression levels of *ARU* and no consistent trend with respect to expression was found.

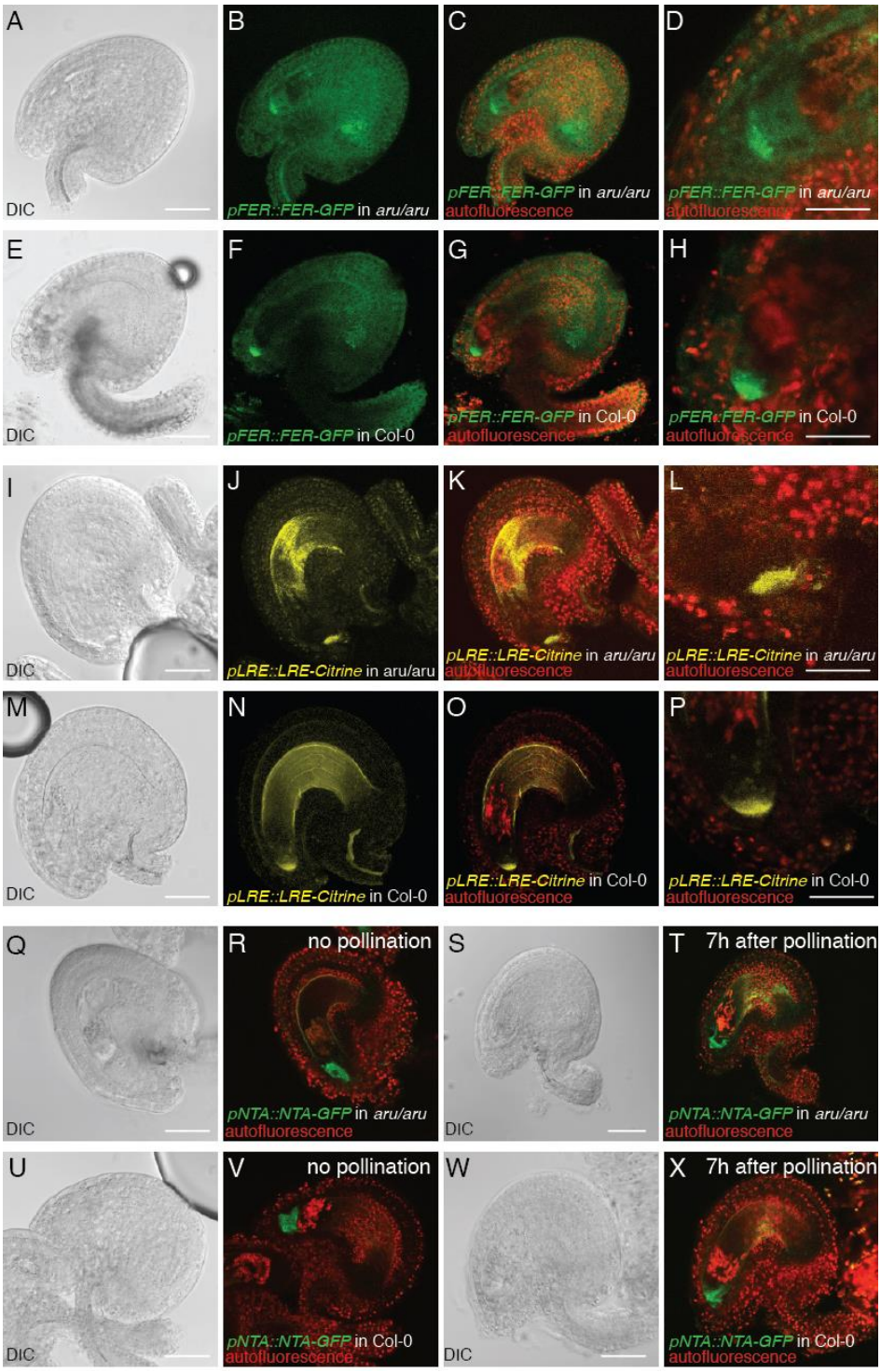


Figure S8: Expression of known factors involved in intraspecific PT reception in *aru* and wild-type ovules. (A-D) *pFER::FER-GFP* in *aru*. (E-F) *pFER::FER-GFP* in Col-0. The strongest FER-GFP expression is detected at the filiform apparatus in synergids both in *aru* and wild-type ovules. (I-L) *pLRE::LRE-Citrine* in *aru*. (M-P*) pLRE::LRE-Citrine* in Col-0. LRE-Citrine is localized exclusively to the filiform apparatus both in *aru* and wild-type embryo sacs. (Q and R) *pNTA::NTA-GFP* in *aru* embryo sacs of emasculated pistils. (S and T) *pNTA::NTA-GFP* in *aru* embryo sacs that received a PT (7 hours after pollination). NTA-GFP is found throughout the synergids before fertilization and is re-localized to the micropylar pole upon fertilization. (U and V) *pNTA::NTA-GFP* in Col-0 ovules of emasculated pistils. (W and X) *pNTA::NTA-GFP* in wild-type ovules that received a PT. There is no visible difference to NTA-GFP in *aru*. Scale bars: 50 µm for all images except (D), (H), (L), (P): 25 µm.

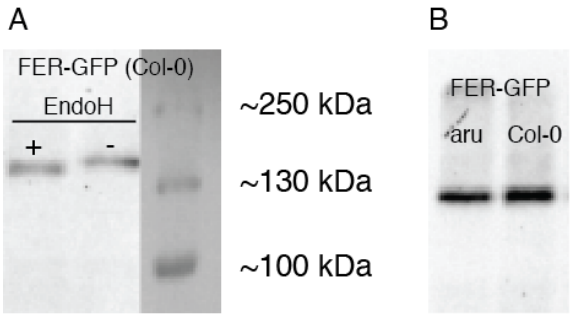


Figure S9: Western blot of FER-GFP in *aru* and wild-type inflorescences. (A) Western blot against FER-GFP in Col-0 background. The sample that was enzymatically deglycosylated with EndoH (+) was running faster on the SDS gel due to lower molecular weight than the untreated sample (-), indicating that FER-GFP is glycosylated *in vivo*. (B) Western blot against FER-GFP in *aru* and Col-0 inflorescences. There are no obvious differences in protein levels and no band-shift visible, which would be caused by altered molecular weight due to differential glycosylation. This indicates that there is no gross change in the glycosylation pattern of FER but does not exclude mis- or underglycosylation.

Table S1: *A. thaliana* accessions used in this study. This table shows their origin (*29*), the blocks they were assigned to, and the mean values of ovules with PT overgrowth per silique (OG/S) that were used for the calculation of the GWAS statistics.

Accession	Region of Origin	Block	Mean values of OG/S (%)
Ag-0	France	Early A, C, D	76.69
An-1	Belgium	Early A, C, D	50.35
Bay-0	Germany	Early B, C, D	57.04
Bil-7	Sweden (North)	Late A, B, C	73.87
Bor-1	Czech Republic	Early A, B, D	46.30
Bor-4	Czech Republic	Early B, C, D	75.56
Br-0	Czech Republic	Early A, C, D	58.51
Bur-0	Ireland	Mid A, B, C	68.08
C24	Portugal	Early B, C, D	39.46
CIBC-17	England	Mid A, B, C	64.46
CIBC-5	England	Early B, C, D	57.62
Col-0	USA (Germany?)	Early A, B, C	65.77
Ct-1	Italy	Early A, B, D	52.68
Cvi-0	Cape Verde Islands	Early A, C, D	75.05
Eden-2	Sweden (North)	Late A, B, C	61.40
Ei-2	Germany	Mid A, B, C	69.39
Est-1	Estland	Early A, B, C	68.34
Fab-2	Sweden (North)	Late A, B, C	73.58
Fab-4	Sweden (North)	Late A, B, C	64.22
Fei-0	Portugal	Early A, B, D	78.14
Ga-0	Germany	Early A, B, C	18.25
Got-22	Germany	Late A, B, C	88.58
Got-7	Germany	Late A, B, C	79.91
Gu-0	Germany	Early A, B, D	60.46
Gy-0	France	Early A, C, D	55.09
HR-10	England	Early B, C, D	75.41
HR-5	England	Early B, C, D	66.29
Kas-1	India	Mid A, B, C	14.81
Kin-0	USA	Early A, C, D	61.82
Knox-10	USA (Indiana)	Mid A, B, C	41.97
Knox-18	USA (Indiana)	Mid A, B, C	59.68
Kondara	Tajikistan	Early A, B, D	56.19
Kz-1	Kazakhstan	Early A, C, D	81.16
Kz-9	Kazakhstan	Early A, B, C	87.30
Ler-1	Poland	Early B, C, D	86.97
Lov-1	Sweden (North)	Late A, B, C	59.19
Lov-5	Sweden (North)	Late A, B, C	46.48
Lp2-2	Czech Republic	Early B, C, D	76.80
Lp2-6	Czech Republic	Early A, B, D	38.89
Lz-0	France	Early A, B, D	10.20
Mr-0	Italy	Late A, B, C	90.93
Mrk-0	Germany	Mid A, B, C	38.10

Mt-0	Libya	Early A, B, D	60.61
Mz-0	Germany	Early B, C, D	70.95
N3	Russia	Mid A, B, C	63.47
Nd-1	Germany	Early A, B, C	88.75
NFA-10	England	Early A, B, C	54.95
NFA-8	England	Early B, C, D	38.30
Nok-3	Netherlands	Mid A, B, C	69.64
Omo2-3	Sweden (South)	Mid A, B, C	58.47
Oy-0	Norway	Mid A, B, C	59.90
Pna-10	USA (Michigan)	Mid A, B, C	33.74
Pna-17	USA (Michigan)	Mid A, B, C	54.57
Pro-0	Spain	Early A, B, D	62.54
Pu2-23	Croatia	Early A, C, D	57.39
Pu2-7	Croatia	Mid A, B, C	65.38
Ra-0	France	Early A, C, D	78.02
Ren-1	France	Mid A, B, C	59.34
Ren-11	France	Early A, B, D	70.88
Rmx-A02	USA (Michigan)	Early A, B, C	44.97
Rmx-A180	USA (Michigan)	Early B, C, D	38.71
RRS-10	USA (Indiana)	Late A, B, C	47.03
RRS-7	USA (Indiana)	Mid A, B, C	45.47
Se-0	Spain	Early A, C, D	68.35
Sha	Tajikistan	Early A, B, D	36.90
Sorbo	Tajikistan	Early A, B, C	46.70
Sq-1	England	Early A, C, D	54.45
Sq-8	England	Early A, B, C	83.45
Tamm-2	Finland	Mid A, B, C	44.85
Tamm-27	Finland	Mid A, B, C	43.77
Ts-1	Spain	Early A, B, D	70.22
Ts-5	Spain	Early B, C, D	63.92
Tsu-1	Japan	Mid A, B, C	62.53
Ull2-3	Sweden (South)	Early A, C, D	64.16
Uod-1	Austria	Early A, B, C	61.70
Uod-7	Austria	Early A, B, C	68.19
Van-0	Canada	Early A, B, D	60.37
Var2-1	Sweden (South)	Late A, B, C	68.16
Wa-1	Poland	Early A, C, D	48.68
Wei-0	Switzerland	Early A, B, C	69.10
Ws-0	Ukraine	Mid A, B, C	56.44
Ws-2	Ukraine	Early A, B, C	37.04
Wt-5	Germany	Mid A, B, C	64.56
Yo-0	USA	Mid A, B, C	51.50
Zdr-1	Czech Republic	Early A, B, D	38.38
Zdr-6	Czech Republic	Early B, C, D	52.47

Table S2: Top40 SNPs of GWAS (TASSEL GLM, (*21*)) ordered by their p-value

Rank	Marker	Chromosome	Position	p
1	PERL0219430	1	24325910	3.80E-07
2	PERL0198933	1	22842107	4.85E-07
3	PERL0998055	5	15096264	2.22E-06
4	PERL0198560	1	22816028	2.24E-06
5	PERL0198559	1	22815992	4.21E-06
6	PERL0667236	4	972178	5.61E-06
7	PERL0661920	4	505633	5.63E-06
8	PERL0062096	1	8320546	6.49E-06
9	PERL0198834	1	22836745	7.19E-06
10	PERL0198872	1	22839590	8.05E-06
11	PERL1016861	5	16420039	1.18E-05
12	PERL0166252	1	19503734	1.31E-05
13	PERL0198952	1	22842689	1.34E-05
14	PERL0329435	2	6392760	1.36E-05
15	PERL0198924	1	22841729	1.38E-05
16	PERL0213344	1	23898307	1.55E-05
17	PERL0198538	1	22814316	1.56E-05
18	PERL0907143	5	6679400	1.65E-05
19	PERL0667367	4	978836	1.86E-05
20	PERL0339587	2	7531633	2.19E-05
21	PERL0339588	2	7531688	2.19E-05
22	PERL1101530	5	23812400	2.22E-05
23	PERL1101540	5	23813337	2.22E-05
24	PERL1101550	5	23813943	2.22E-05
25	PERL1101566	5	23815590	2.22E-05
26	PERL1101568	5	23815756	2.22E-05
27	PERL1101569	5	23815796	2.22E-05
28	PERL0291095	2	2220611	2.53E-05
29	PERL0617038	3	18010190	2.58E-05
30	PERL0667057	4	957961	2.72E-05
31	PERL0978994	5	13673830	2.74E-05
32	PERL0277850	2	1030731	3.07E-05
33	PERL1015945	5	16345246	3.14E-05
34	PERL0996614	5	15041429	3.18E-05
35	PERL0219784	1	24354156	3.26E-05
36	PERL0726223	4	6441807	3.31E-05
37	PERL0217577	1	24223030	3.33E-05
38	PERL0233274	1	25610774	3.53E-05
39	PERL0891205	5	4924090	3.57E-05
40	PERL1101497	5	23809336	4.00E-05

Table S3: Mutant lines used in this study.

Affected gene	**Mutant allele**
At1g61790 (*aru-1*)	SALK_067271
At1g61790 (*aru-2*)	EMS allele (Peter McCourt)
At1g61780	SALK_137883C
At1g61795	SALK_052207C
At1g61795	SALK_026074C
At1g61810	SALK_104077

Additional Results for Chapter 3

These experiments are not included in the manuscript but will still be presented in this thesis.

Additional Results for Chapter 3

Association mapping using mixed models

Mixed linear models (MLM) allow correction for the strong population structure present in the worldwide population of *A. thaliana* (*1*). In contrast, general linear models (GLM) constitute a simplified analysis that does not take population structure into account. GLM corresponds to MLM with maximal compression, increasing false positive rate (*2*). In turn, MLM are prone to increase false negatives (true interactions are masked). In addition to GLM (chapter 3), we used several MLM algorithms, including GAPIT and GWAPP (described in the main section of this chapter), as well as EMMA, EMMAX and the MLM P3D-function implemented in TASSEL (*2-7*). EMMA (emma.REML.t option), EMMAX and P3D gave similar results and thus, only the TASSEL analysis will be presented here. For the analysis, we used two different SNP datasets, the full set (215983 SNPs, (*9*)) and a reduced dataset containing only SNPs with an allele frequency > 0.1 (SNP0.1; 179481 SNPs). This enables us to discard very rare alleles that might skew the analysis. The 5% Bonferroni-corrected p-value is $3 \times e^{-7}$ for the SNP01 dataset and $2 \times e^{-7}$ for the full SNP set, and the 10% values are $6 \times e^{-7}$ and $5 \times e^{-7}$, respectively. No clear or significant peaks were observed and highly (yet non-significant) correlated SNPs seemed to be singletons (neighbouring SNPs not linked) (Figure 1A, B). For the Top 100 SNPs with the lowest p-values see Appendix (tables A1 and A2).

Although no significant peaks could be detected using MLM, we analyzed candidate mutants of genes that are harboring or are in close proximity of SNPs close to significance, within a clear peak (neighboring SNPs are linked, too) or coming up in several MLM methods. Candidate mutants were pollinated with *A. lyrata* pollen and overgrowth was assessed two days after pollination after Aniline Blue staining. None of the candidate mutants showed an interspecific OG/S that clearly differed from wild type *A. thaliana* (Col-0) pollinated with *A. lyrata* (Fig. ad3.1 C-E, table ad3.1).

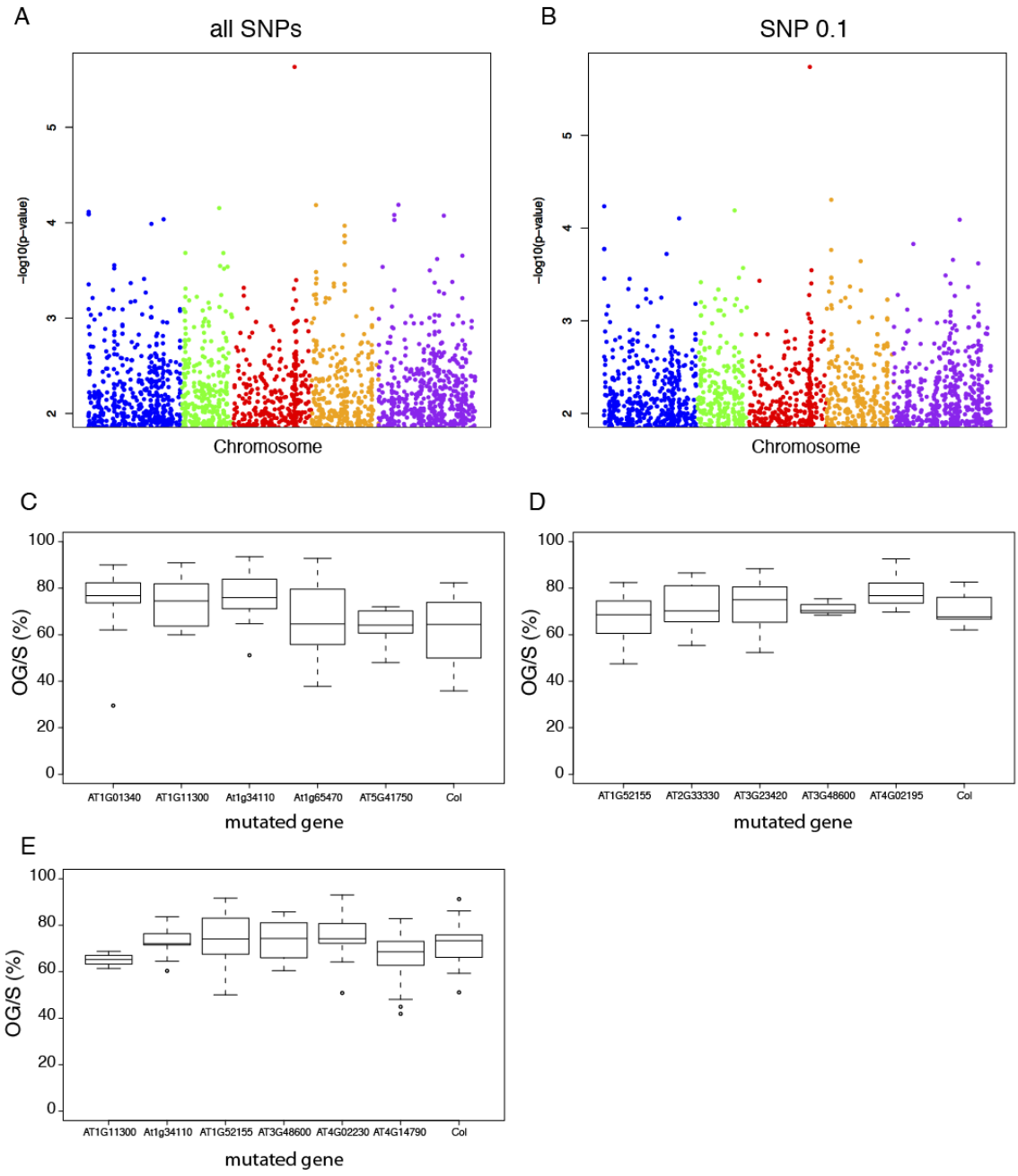


Figure ad3.1: GWAS analysis for OG/S using P3D and mutant analysis in interspecific crosses. (**A+B**) Manhattan plots of GWAS results using the P3D algorithm implemented in TASSEL (*2*, *7*). All correlated SNPs are below the 5%-Bonferroni-corrected significance threshold (not shown in graph; $-\log_{10}(2 \times e^{-7})=6.7$ for the full SNP set, $-\log_{10}(3 \times e^{-7})=6.52$ for the SNP0.1 Set) and most of them are singletons. Colors represent the five chromosomes. (A) Manhattan plot using the full SNP set. (B) Manhattan plot with SNP0.1 set. (**C-E**) Proportions of ovules with *A. lyrata* PT overgrowth per silique (OG/S) in candidate mutants derived from GWAS MLM analyses. No obvious phenotype of one of the mutants compared to Col-0 is visible. (C), (D) and (E) are three different experiments (batches) conducted at three different days and are therefore plotted separately.

Table ad3.1
Mutant lines used for interspecific pollinations and their assignment to three batches (compare Fig. 1)

Gene	Function	T-DNA Lines	Batch
AT1G01340	Member of cyclic nucleotide gated channel family	SALK_071112C	I
		SALK_096435	
AT1G11300	Protein serine/threonine kinase	SALK_044069	I, III
		SAIL_150_H02	
At1g34110	Leucine-rich receptor-like protein kinase family protein	SALK_058918	I, III
		SALK_039687	
AT1G52155	unknown protein	SALK_136633C	II, III
		SALK_043049	
At1g65470	Chromatin Assembly Factor-1 (CAF-1) p150 subunit	*fas1-1*	I
		SALK_009476	
AT2G33330	Plasmodesmatal protein	SALK_145882	II
		SALK_016528	
AT3G23420	Transcription factor INNER NO OUTER	SALK_116219	II
AT3G48600	SWIB complex BAF60b domain-containing protein	SALK_008134	II
		WiscDsLox402G01	
AT4G02195	SYB4 family member	SALK_116966C	II
		SALK_126693	
AT4G02230	Ribosomal protein L19e family	SALK_042253	III
		SALK_082133	
AT4G14790	DExH box RNA helicase	SALK_078210	III
		SALK_029597	
AT5G41750	Disease resistance protein	SALK_133292	I
		SAIL_861_H05	

In addition, we conducted GWAS using a Multi-Locus Mixed Model (MLMM) (*8*). This model assumes that a trait is determined by more than one locus and tests for multiple, dependent SNPs in a stepwise analysis. Using MLMM, we identified a strong peak on chromosome 5 in step 5, but only analyzed a mutant line of one candidate gene (*At5g62230* coding for the receptor-like kinase *ERECTA-LIKE 1*, SALK_084012C) in interspecific crosses and did not see a phenotype different from wild type. However, this region on chromosome 5 seems to be crucial for interspecific PT reception as it was also identified using a second mapping approach (see chapter 4 for details, the MLMM plot and the mutant phenotype).

Frequency analysis of SNPs surrounding *ARU*

In the SNP dataset we used for our GWAS GLM analysis, only SNPs that surround the coding region of *ARU* were identified. Of the three SNPs that show the highest correlation to interspecific PT reception around *ARU*, two are 562bp and 598bp downstream of the coding region, respectively (Positions 22815992, PERL0198559 and 22816028, PERL0198560), and one is 74bp upstream of the translation start (Position 22814316, PERL0198538). This polymorphism segregates 64%C/36% A in the accessions used for the phenotyping, with the A allele being more frequent in accessions with low PT overgrowth in interspecific crosses (Fig. ad3.3). The SNPs downstream segregate 63%G/37%C (PERL0198559) and 64%G/36%C (PERL0198560), respectively. In both cases, the lower-frequency C allele is associated with low PT overgrowth.

The *ARU* allele from Ga-0 cannot complement *aru*

In order to analyze if the *ARU* allele from a low OG/S accessions (Ga-0) can complement the *aru* mutant better than the allele from an intermediate one (Col-0), we created transgenic *aru* lines expressing the Ga-0 allele of *ARU* under the control of the endogenous promoter (including 1492 bp upstream of the ATG) with additionally 865 bp of downstream sequence after of the stop-codon. This fragment includes all of the above-mentioned SNPs in the GWAS SNP dataset; and for all three polymorphisms, Ga-0 carries the allele associated with low interspecific PT overgrowth. In addition, several other SNPs compared to Col-0 were detected after sequencing the fragment, one of which leads to an amino acid exchange in the coding region (compare chapter 3). These are upstream SNPs: -1270 bp: T→G, -1037 bp: A→C, -788 bp: T→C, -456 bp: C→T; SNPs within the gene at position 198: G→T and position 976: G→A (Val→Ile); and downstream SNPs: +280: T→G, +412/413: TT→GA (– and + indicate SNP positions upstream of the ATG and downstream of the stop-codon, respectively. The first base is the Col-0 reference sequence, and the second one is the respective base in Ga-0). In addition, the same fragment derived from Col-0 was cloned as a control. For both constructs, the genomic *ARU* fragment was amplified using primers 5' – TTTTACTAGTAGGCAATTCCATCAGTTGTT – 3' and 5' – TTTTGGTACCGTTACTTCACTTTCTCGAGT – 3', introducing a *Spe*I and a *Kpn*I restriction

site, respectively. The fragments were cloned into pMDC99 using restriction-ligation and transformed into *aru* mutants (Col-0 background).

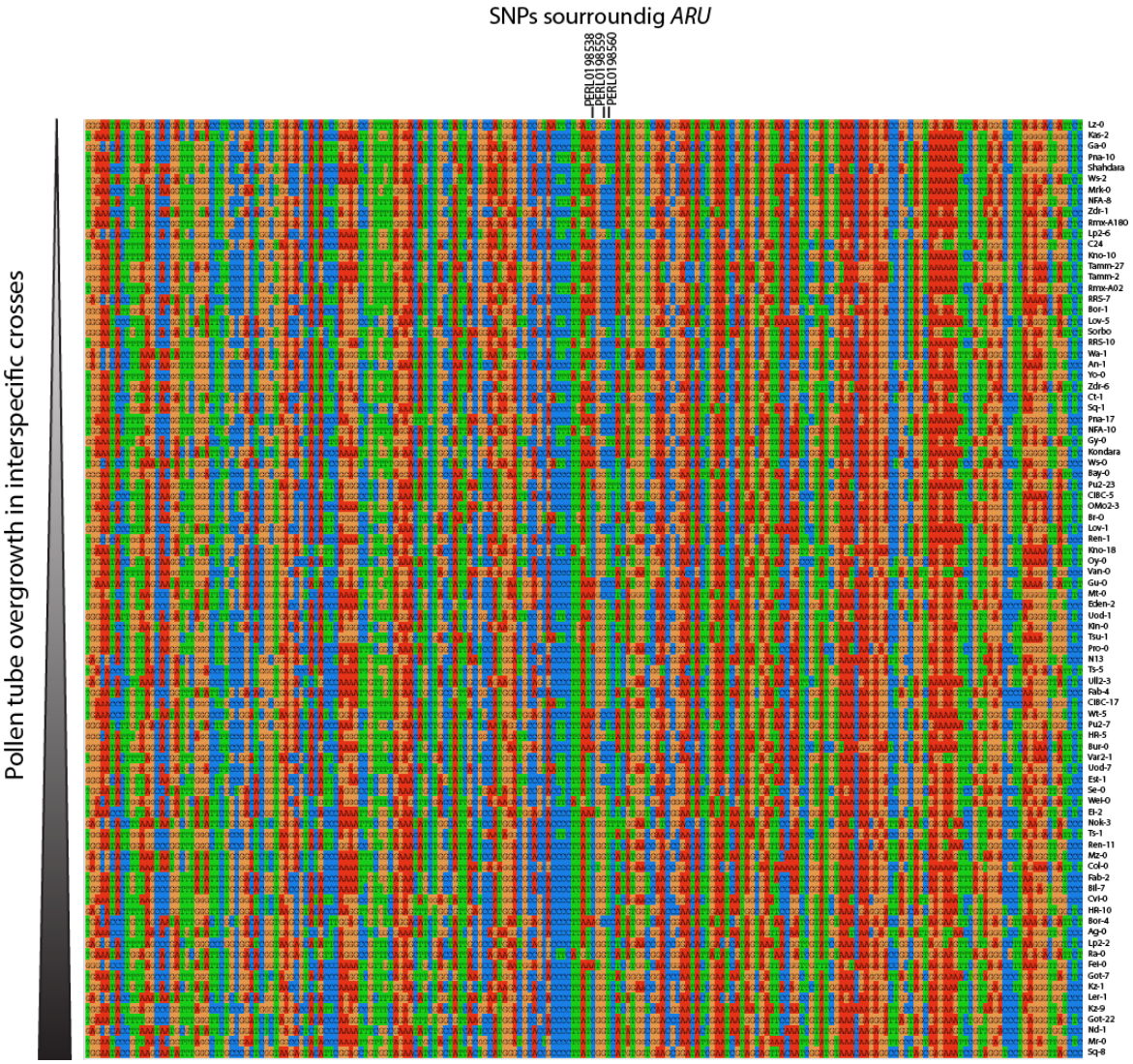


Figure ad3.3: Alignment SNPs in the proximity of ARU and their correlation to the OG/S phenotype. Accessions are ordered by their degree of interspecific PT rejection (top: low OG/S, bottom: high OG/S) and SNPs in the vicinity of *ARU* are marked with their PERL identifiers.

In order to determine if the Ga-0 version of *ARU* is able to complement the *aru* mutant better than the Col-0 allele (less ovules with interspecific PT overgrowth, OG/S), we pollinated homozygous transgenic plants (T3 generation) with *A. lyrata* pollen 2 days after emasculation and assessed PT overgrowth. In transgenic plants carrying the Ga-0 allele, a very low percentage of PT overgrowth per silique would indicate that *ARU* (Ga-

0) alone is able to confer the big phenotypic difference in interspecific PT overgrowth observed between Ga-0 and Col-0 (18.25 and 65.77%, respectively).

Unfortunately, we were only able to recover one transgenic plant carrying the Ga-0 allele and this line does not seem to complement *aru* at all (Fig. ad3.4 A). In contrast, the Col-0 allele rescues the mutant phenotype in various degrees, indicating that the promoter is functional (compare also the ARU-GFP synergid localization when expressed under the same 1492 bp endogenous promoter; chapter 3). The fact that the Ga-0 allele does not seem to complement the mutant phenotype at all could be because we used only one transgenic line carrying this allele, which by chance might have not been expressed. Another possibility is that in Ga-0, additional trans-regulatory factors are required for the cell-type specific expression of *ARU*, and that those factors are not present in the *aru* mutant (Col-0 background). Furthermore, it is possible that we did not include all necessary cis-regulatory elements required for Ga-0 *ARU* expression. The analysis of more lines carrying the *ARU* (Ga-0) transgene would help to solve this question. However, results from our bulk-segregant analysis (see chapter 4 of this thesis) indicate that *ARU* is not responsible for the creation of the big phenotypic difference between Ga-0 and high OG/S accessions like Fei-0 or Sq-8 (78.14 and 83.45% OG/S). It is thus conceivable that *ARU* is only gradually mediating the phenotypic variation observed in the 86 accessions, rather than causing the big variation between extreme accessions alone.

The homolog of *ARU* does not show a phenotype in PT reception

The homolog of *ARU* in *A. thaliana* is called *OST3/6-LIKE* and encoded by *At1g11560* (*11*). This gene is not expressed in synergids and also absent in all other reproductive tissues (egg and central cell as well as pollen and sperm cells) according to microarray data (*12*). In order to analyze the effect of *OST3/6-LIKE* in interspecific PT reception, we pollinated mutants with *A. lyrata* pollen and analyzed the proportion of ovules with PT overgrowth in the siliques. As expected for a gene not expressed in the synergid cells mediating PT reception, *ost3/6-like* mutants (GK-044G02) did not show differences in OG/S compared to wild type (mean: 56.63% in *ost3/6-like* and 60.24% in Col-0; Fig ad3.4 B).

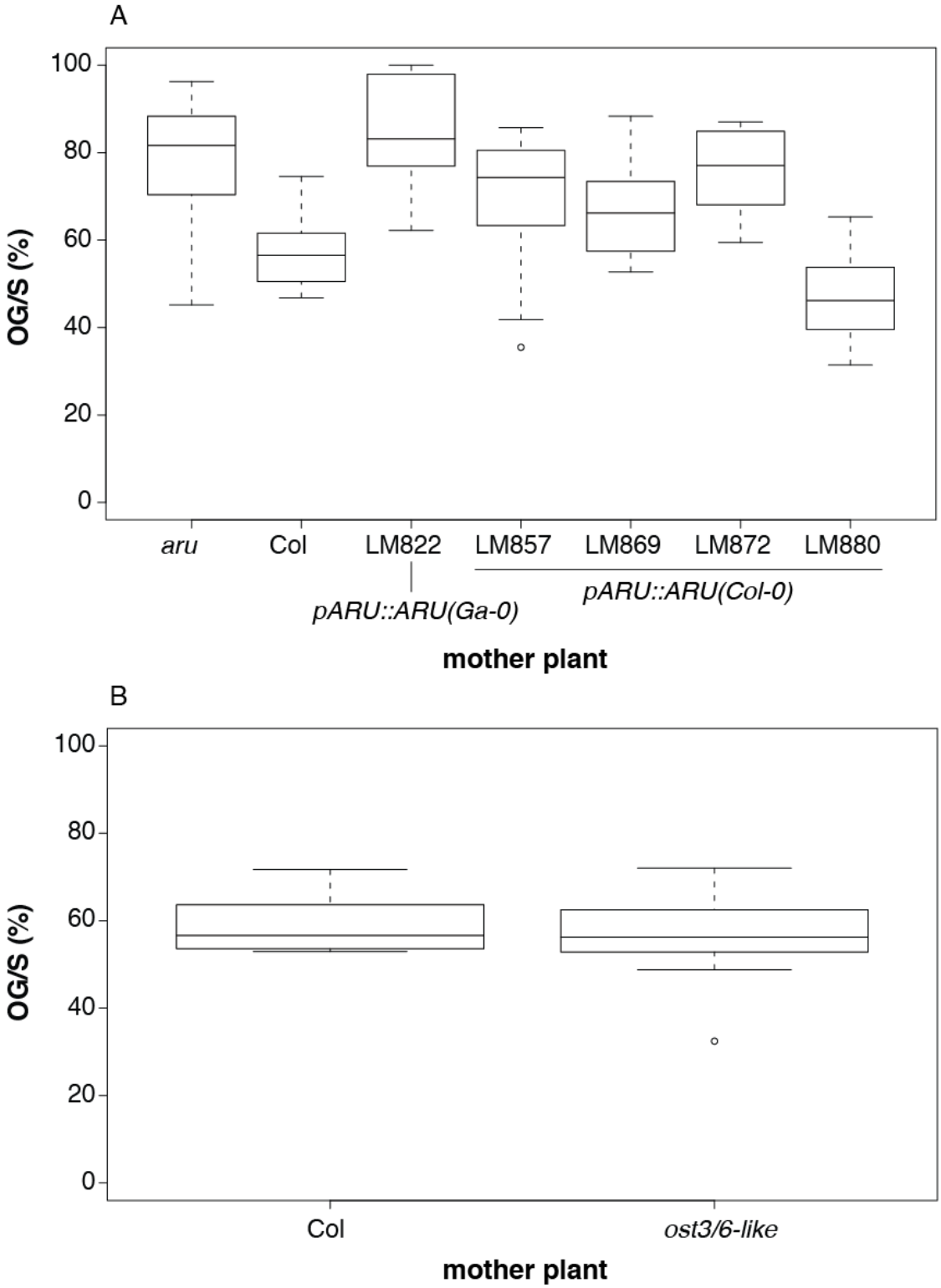


Figure ad3.4: Interspecific pollination of complemented *aru* as well as *ost3/6-like* mutants. (**A**) *aru* mutants expressing *ARU* from Ga-0 or Col-0 pollinated with *A. lyrata*. One line with the Ga-0 allele does not complement the mutant phenotype, whereas the lines with the Col-0 allele do. (**B**) Mutants in the *ARU*-homolog, *ost3/6-like*, pollinated with *A. lyrata* pollen. No difference in OG/S compared to wild type is visible.

The production of viable hybrid seeds is accession-dependent

For our GWAS phenotyping, pollinated siliques were harvested two days after pollination. Ovules with normal PT reception in interspecific crosses (no overgrowth) appeared to be fertilized by *A. lyrata* PTs because they seemed to develop normal endosperm. In order to further examine this, we allowed few siliques of crosses of selected accessions (the low OG/S accessions Lz-0, Kas-1, Ga-0, Lp2-6, and Col-0) with *A. lyrata* to develop seeds and analyzed the progeny. Interestingly, the crosses of Kas-1 and Lp2-6 with *A. lyrata* pollen did not yield in viable hybrid seeds (shrunken seeds of Kas-1 x *A. lyrata* hybrids and no recovery of viable seedlings of Lp2-6 x *A. lyrata* F_1). For the other accessions, hybrid, sterile-looking F_1 plants could be recovered, indicating that in Kas-1 and Lp2-6, post-zygotic hybridization barriers must act that are absent in Lz-0, Ga-0 and Col-0. It would be interesting to determine the nature of these barriers, possibly by using a GWAS approach.

The hybrid F_1 plants were pollinated with both *A. lyrata* and *A. thaliana* (Col-0) pollen to recuperate seeds, however, this was not successful. Pollen of both species adhered to the stigma and PTs grew towards the transmitting tract, but were not attracted towards the ovules. This is in contrast to published work, which reports that F_1 backcrosses to both parental species (*A. lyrata* and *A. thaliana* Col-0) lead to the production of viable seeds (*13*). However, we only looked at small numbers of siliques and it is conceivable that we would have also succeeded in generating backcross progeny by doing more crosses.

Phenotypic variation of interspecific PT attraction

For the trait „interspecific PT overgrowth per silique“, only ovules that had received a PT (functional PT guidance) were taken into account. However, during counting, we simultaneously assessed the number of ovules that did not attract an interspecific PT. Less variation than in the PT overgrowth trait was observed between accessions, but the broad-sense heritability H^2 was in a similar range (74.5%; Fig. ad3.5 A). Attempts do conduct GWAS analysis on the mean values of ovules that had not attracted an interspecific PT resulted in several peaks and singletons using EMMA MLM (for a definition of the full and the SNP0.1 dataset see above) but were not pursued further

(Fig. ad3.5 B, C). See Appendix for a list of the Top 100 SNPs with the highest correlation to the trait (tables A3 and A4).

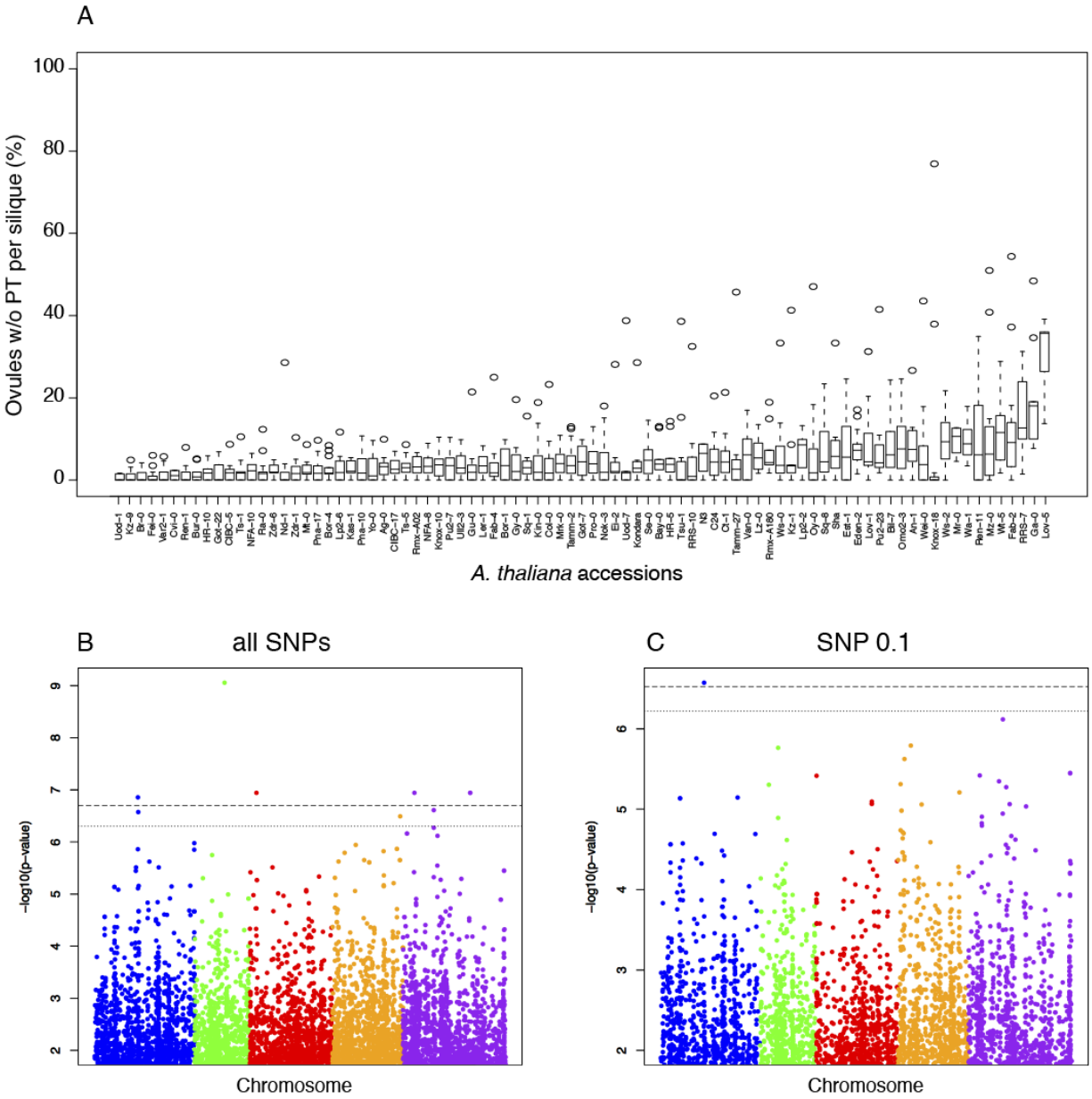


Figure ad3.5: Natural variation in PT attraction in 86 accessions. (**A**) Proportions of ovules in a silique that did not attract *A. lyrata* PTs. (**B+C**) GWAS EMMA analysis on that trait using the full SNP set (B) and the SNP0.1 set (C). Dashed line: 5% significance threshold, dotted line: 10% significance threshold (Bonferroni-corrected).

References

1. M. Nordborg *et al.*, The pattern of polymorphism in *Arabidopsis thaliana*, *PLoS biology* **3**, (2005).

2. Z. Zhang *et al.*, Mixed linear model approach adapted for genome-wide association studies, *Nature genetics* **42**, 355–360 (2010).

3. A. E. Lipka *et al.*, GAPIT: genome association and prediction integrated tool, *Bioinformatics* **28**, 2397–2399 (2012).

4. U. Seren *et al.*, GWAPP: a web application for genome-wide association mapping in *Arabidopsis*, *The Plant cell* **24**, 4793–4805 (2012).

5. H. M. Kang *et al.*, Efficient control of population structure in model organism association mapping, *Genetics* **178**, 1709–1723 (2008).

6. H. M. Kang *et al.*, Variance component model to account for sample structure in genome-wide association studies, *Nature genetics* **42**, 348–354 (2010).

7. P. J. Bradbury *et al.*, TASSEL: software for association mapping of complex traits in diverse samples, *Bioinformatics* **23**, 2633–2635 (2007).

8. V. Segura *et al.*, An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, *Nature genetics* **44**, 825–830 (2012).

9. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, *Nature* **465**, 627–631 (2010).

10. J. M. Escobar-Restrepo *et al.*, The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception, *Science* **317**, 656–660 (2007).

11. A. Farid *et al.*, Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses, *Plant physiology* **162**, 24–38 (2013).

12. S. E. Wuest *et al.*, *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes, *Current biology* **20**, 506–512 (2010).

13. M. E. Nasrallah, K. Yogeeswaran, S. Snyder, J. B. Nasrallah, *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants, *Plant physiology* **124**, 1605–1614 (2000).

Chapter 4 – Results

A small LysM-peptide mediates interspecific pollen tube reception

This chapter constitutes a basis for a manuscript that will be submitted after additional experiments (mentioned in the text) have been conducted.

A small LysM-peptide mediates interspecific pollen tube reception

Lena M. Müller, Nuno D. Pires, Marc W. Schmid, Ueli Grossniklaus

Abstract

During plant reproduction, pollen tube (PT) reception is a cell-cell communication process allowing the recognition of the male gametophyte (the PT) by the female gametophyte (the embryo sac). Upon successful recognition, the PT bursts and releases its sperm cells in order to effect double fertilization of the female egg and central cell. In interspecific crosses between certain *A. thaliana* accessions and *A. lyrata* pollen, most PTs are not recognized, causing a failure in PT growth arrest and thus leading to PT overgrowth inside the female gametophytes. In contrast, several other *A. thaliana* accessions almost perfectly recognize interspecific PTs. Using a combination of bulk-segregant analysis and reverse genetics, we were able to identify *At5g62150* as a key player for species-specific PT recognition. *At5g62150* codes for a small, LysM-domain containing glycan-binding peptide, which is after the identification of *ARTUMES* once more pointing at the importance of glycosylation patterns for the distinction of intra- and interspecific PTs.

Introduction

In order to ensure healthy offspring, plants have developed a variety of species barriers that prevent the formation of potentially unviable or sterile hybrid seeds. There are several types of hybridization barriers, some act before pollination (pre-pollination barriers) and some after (post-pollination barriers) (*1*). Post-pollination prezygotic barriers prevent the formation of a zygote after pollen from a different species has landed on a plant's stigma and rely solely on cell-cell communication between the male gametophyte and the female tissues such as stigma, transmitting tract and the embryo sacs. After the transfer of pollen onto a stigma, it hydrates and germinates an elongating, tip-growing cell (the PT), which transports the male sperm cells through the female tissues towards the ovules. Once arrived, the two gametophytes communicate and upon recognition of the PT by the embryo sac, the PT arrests its growth, bursts and releases

its sperm cells into the female gametophyte (Fig. 4.1B). This process is called PT reception and is to a large extent mediated by the synergid cells which possess a thickened cell wall structure at the micropylar end (the filiform apparatus), which is believed to account for most of the communication (*2*). A failure in recognition of the PT by the female gametophyte leads to continued growth of the PT inside the embryo sacs without bursting, leaving the female gametophytes unfertilized (Fig. 4.1A). This phenomenon is known as PT overgrowth and has been observed in a variety of male or female gametophytic mutants, as well as in interspecific crosses within the families Ericaceae and Brassicaceae (*3*-*9*). Recently, we showed that recognition of several interspecific *Arabidopsis* PTs by *A. thaliana* is dependent on the accession of the mother plant and that the communication between *A. lyrata* PTs and *A. thaliana* embryo sacs is dependent on glycosylation of unknown target proteins by the oligosaccharyltransferase subunit *ARTUMES* (*ARU*), which seems to confer species-specificity (Chapter 3 of this thesis). We have identified *ARU* with the help of a genome-wide association study (GWAS) based on the proportion of ovules with PT overgrowth per silique (OG/S) in 86 *A. thaliana* accessions pollinated with *A. lyrata* pollen. The phenotypic variation among the accessions ranged from about 10% to more than 90% OG/S and displayed a gradual distribution suggesting multiple genes to be involved in the specification of the trait (Chapter 3). In order to identify more players in the recognition of interspecific PTs, we created mapping populations of accessions with extreme OG/S phenotypes and identified *At5g62150* as a regulator of interspecific PT reception by a combination of bulk-segregant analysis (BSA) and reverse genetics.

Results

Generation of mapping populations based on accessions with extreme OG/S

Previously, we have shown that *A. thaliana* accessions differ in their ability to recognize and receive PTs from closely related *Arabidopsis* species. In siliques of accessions like Lz-0 and Ga-0, only about 10.2% and 18.3% of ovules reject *A. lyrata* PTs (characterized by interspecific PT overgrowth), whereas the majority of the ovules accept the interspecific PTs (Chapter 3). In contrast, accessions like Fei-0 and Sq-8 have much higher values of OG/S (78.1% and 83.4% in the original experiment, chapter 3). In order to make use of this natural variation and to determine the underlying genetic basis causing the phenotypic differences between pairs of accessions, we crossed accessions

with extreme low and high OG/S values to generate a mapping population. In total, we used three accessions with low OG/S (further denoted as "low" accessions: the above-mentioned accessions Lz-0 and Ga-0 as well as Kas-1, which had an overgrowth rate of 14.8% in the original experiment). These low accessions were crossed in all possible combinations as pollen donors to five accessions that displayed high OG/S ("high" accessions: Fei-0 and Sq-8 as well as Kz-1 with 81.1%, Kz-9 with 87.3%, and Nd-1 with 88.8% OG/S). The only exceptions were Kas-1 x Nd-1 and Lz-0 x Nd-1, where the low parent was used as a mother plant to create the hybrids. The parental accessions were chosen by their extreme phenotype and by the relatively uniform time they require for flowering (*10*).

F_1 of accession hybrids display intermediate OG/S compared to the parental accessions

First, we assessed the phenotype of the hybrid F_1 generation resulting from the various accession combinations in interspecific crosses with *A. lyrata*. Most F_1 displayed a proportion of PT overgrowth intermediate between the OG/S rates of their parental accessions (Fig. 4.1, fig. S1). This suggests that the variation in interspecific PT reception between the accessions is either gametophytically regulated, or that a pair of alleles derived from the parental accessions is acting in sporophytic pistil tissues contributing additively to the intermediate phenotype in the F_1. For example, F_1 hybrids between Ga-0 (14.8% OG/S) and Sq-8 (86.4% OG/S) display 52.4% ovules with PT overgrowth per silique (Fig. 4.1C). However, F_1 hybrids originating from a cross of Nd-1 or Sq-8 with Kas-1 deviated from this general observation. Those F_1 hybrids showed proportions of PT overgrowth resembling the high parental accession (fig. S1). Thus, it is conceivable that in combinations between Kas-1 and Nd-1 or Sq-8, one or more genes acting dominantly in sporophytic tissues are involved in establishing the phenotypic variation between the parental accessions.

Segregation of F_2 populations depends on the combination of parental accessions

In order to identify the number of genes determining the variation in OG/S between accession pairs, we analyzed the segregation of phenotypes in F_2 populations of selected parental combinations (derived from the original crosses Fei-0 x Ga-0, Fei-0 x Lz-0, Kz-1

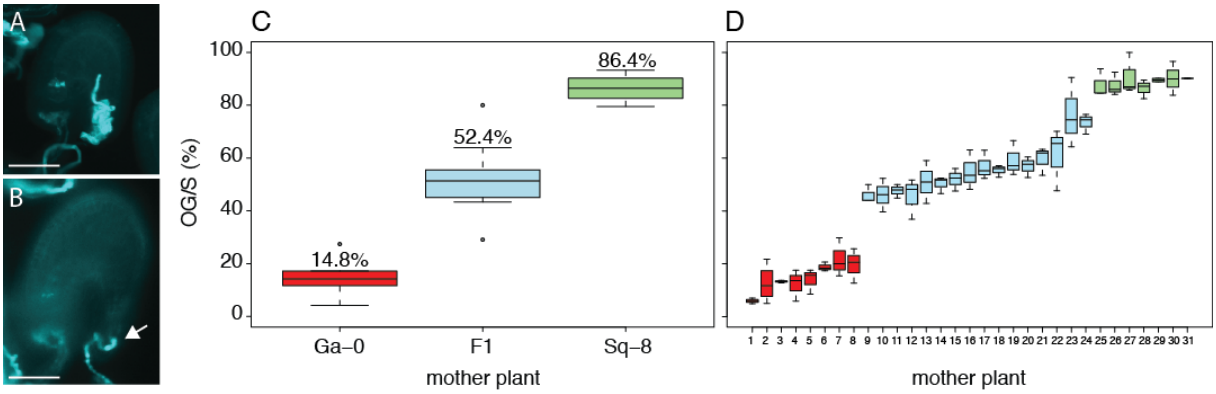


Figure 4.1: Interspecific PT overgrowth is dependent on the genotype of the *A. thaliana* mother plant. (**A**) An *A. thaliana* ovule with an interspecific *A. lyrata* PT overgrowing in the female gametophyte. This ovule remains unfertilized. (**B**) An *A. thaliana* ovule, which has been fertilized by an interspecific PT. Remainders of the ruptured PT are visible at the micropylar end (arrow). (**C+D**) Ovules with interspecific *A. lyrata* PT overgrowth per silique (OG/S). (C) The F_1 (blue) of the initial cross Ga-0 (red, low OG/S) and Sq-8 (green, high OG/S) shows an intermediate OG/S phenotype compared to the parents. Numbers on top of the boxplots denote the respective means. (D) Interspecific OG/S of 31 individuals of a 1:2:1 (low:intermediate:high OG/S) segregating F_2 population derived from an Sq-8 x Ga-0 cross pollinated with *A. lyrata*. Individuals with low OG/S phenocopy Ga-0 (red), individuals with high OG/S correspond to the 95% confidence interval of Sq-8 (green). Blue: Individuals with intermediate OG/S levels that could not be assigned to the confidence intervals of any of the parental accessions. Scale bar in (A) and (B): 50µm.

x Lz-0, Kz-9 x Lz-0, Nd-1 x Ga-0, Lz-0 x Nd-1, Sq-8 x Ga-0 and Sq-8 x Lz-0; table S1). Because we were mainly interested in discovering the genetic basis of interspecific PT reception, we did not include F_2 populations originating from crosses with Kas-1 in our analysis, as we suspected complex genetic interactions in these populations that could complicate the identification of single genes. Three of eight F_2 populations showed a 1:2:1 segregation pattern (low:intermediate:high OG/S), indicating that between the original parental accessions, only a single locus is causing the dramatic phenotypic difference, which is segregating in the F_2. Taken together with the intermediate phenotype of the F_1, it can be concluded that this single gene must act in the female gametophyte and thus we can exclude at least for these F_2 populations the possibility that a pair of sporophytically active loci is additively causing the intermediate F_1 phenotype. These three 1:2:1 segregating populations originated from the original crosses Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0 (fig. S2, table S1). For example in the F_2 originating from Sq-8 x Ga-0, 8 individuals showed OG/S ranging from 5.93 – 21.75% resembling Ga-0, 16 intermediate rates (45.6 – 76.4%) and 7 individuals resembled the high Sq-8 parent, displaying overgrowth rates of 86.42 – 90.89% (Fig 4.1D). The

observed high OG/S values are within or above the 95% confidence interval for Sq-8 calculated from our initial results (77,51–89,39%) and also the low values are within or even below the confidence interval for Ga-0 (12,37–24,12%) (table S2, Chapter 3). The observed ratio of 8:16:7 in the Sq-8 x Ga-0 F_2 population (low:intermediate:high) is not different from a 1:2:1 distribution expected for segregation of a single locus (p=0.95; chi-square test goodness of fit). The same is true for the two other populations that segregate 1:2:1 (Fei-0 x Ga-0, 8:14:5, p=0.70; Fei-0 x Lz-0, 8:18:10, p=0.89; table S2). The other analyzed F_2 populations showed continuous OG/S values across individuals, indicating that the phenotype is determined by multiple genes in those populations (quantitative trait, fig. S2).

Bulk-Segregant Analysis identifies a linked region on chromosome 5

Because we had evidence that variation in a single locus seems to be able to cause a phenotypic difference between less than 20% of interspecific PT overgrowth (Ga-0 and Lz-0) and about 80% (Fei-0 and Sq-8), we aimed to identify this key player in interspecific PT recognition. Thus, we screened a larger number of F_2 individuals derived from the original crosses Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0 in interspecific crosses and pooled leaf material of >100 segregating individuals with either extreme high or low OG/S separately for each F_2 population (denoted as “high bulk” and “low bulk”, respectively) (table S3). Individuals with intermediate phenotypes were discarded. DNA extracted from those six pools was sequenced using Illumina HiSeq technology and reads were assembled to “reference genomes” for both respective parental accessions. Those “reference genomes” were created by substituting nucleotide bases in the Col-0 reference sequence with accession-specific SNPs from publicly available sources (Ga-0, Fei-0, and Sq-8 from the 1001 genomes project (*11*); Lz-0 from (*12*)). We expect that in the genomic region responsible for the variation in interspecific PT reception, all reads of the low pool should resemble Ga-0 and Lz-0, respectively, whereas in the high pool, all reads should be attributed to Fei-0 or Sq-0 (resulting in allele frequencies of 1 at the causative locus). In unlinked regions, the parental SNPs should segregate randomly with an allele frequency of 0.5. We calculated the likelihood that a given genomic region is associated with the variation between the low and high pools using a method based on the standard G statistic for Bulk-Segregant Analysis (BSA) originally proposed by (*13*) and (*14*). This statistic was smoothened by averaging

G values using a sliding window of 5000 SNPs (G'). We were able to identify a significantly linked region which spans almost the whole chromosome 5 (False Discovery Rate, FDR <0.01) for each of the three F_2 populations (Fig. 4.2 and fig. S3). However, the G' peaks of the Fei-0 x Ga-0 and the Sq-8 x Ga-0 populations were only 11.7 kbp away from each other on the lower arm of chromosome 5 (G' peaks at position 24995027 for Fei-0 x Ga-0 and at position 24983290 for Sq-8 x Ga-0). Interestingly, the same region was identified with a GWAS multi-locus mixed model (MLMM) using mean OG/S values for all 86 accessions from our previous experiment (peak at position 24996125 in step 5, fig. S4, table S4; (*15*), Chapter 3). The G' peak of the Fei-0 x Lz-0 population is about 1.1 Mbp away from the Fei-0 x Ga-0 and Sq-8 x Ga-0 peaks (positions 23907401-465 with the same G' value). However, it seems that this G' statistic has multiple smaller peaks, one of them again pointing to the same region identified with the other two F_2 populations (positions 24944386-24952426; fig. S4). However, because two BSA peaks as well as the GWAS MLMM peak clearly identified the same chromosomal sector, we denoted a candidate region spanning +/- 50 kb around the Fei-0 x Ga-0 and the Sq-8 x Ga-0 and the minor Fei-0 x Lz-0 peak (table S5).

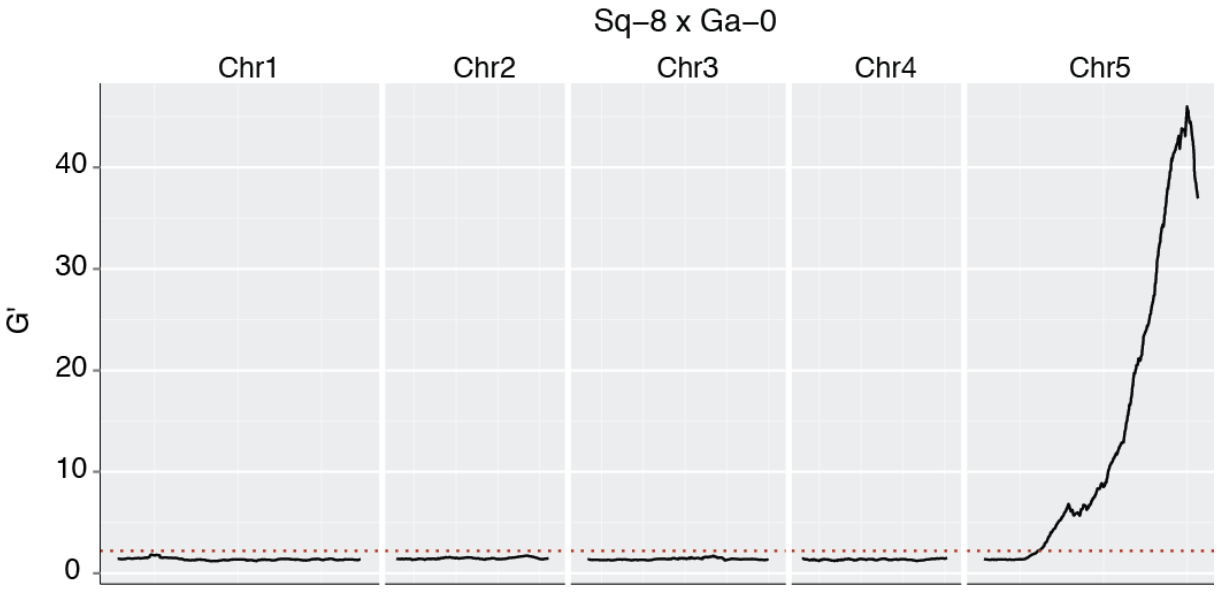


Figure 4.2: BSA reveals a peak on chromosome 5 that causes OG/S variation between Ga-0 and Sq-8. Most of chromosome 5 is significantly linked to the variation of the trait (dotted line: FDR <0.01), but a clear peak is visible (position 24983290).

Mutants in *At5g62150* phenocopy "high" accessions

The defined 150 kb QTL contains 43 genes. In order to narrow down the candidate region to a single gene, we obtained T-DNA insertion lines for all except eight genes in that region and analyzed their phenotype in interspecific crosses with *A. lyrata* (table S5, fig. S5). Among all T-DNA insertion lines tested, a homozygous mutant line in *At5g62150* (SALK_144729) showed a phenotype strongly divergent from the one observed in wild type: in the mutant, an average of 83.29% of ovules in a silique displayed PT overgrowth compared to 58.73% in the Col-0 wild type ($p<0.0001$), thus resembling OG/S of the high accessions Fei-0 and Sq-8 (85.49% and 79.92%, respectively; Fig. 4.3A-D and G, fig. S5). In addition, *At5g62150* is expressed in the synergid cells according to microarray data and therefore constitutes a likely candidate for a mediator of PT reception (*16*). Interestingly, mutants in *At5g62150* do not show differences in intraspecific reception of *A. thaliana* PTs compared to wild type (Fig. 4.3E-F and H, $p=0.4473$), further strengthening its role in the recognition and distinction of intra- and interspecific pollen tubes.

At5g62150 codes for a small peptide with a Lysin Motif

At5g62150 encodes a small (102 amino acids) LysM domain-containing protein with unknown molecular function. The Lysin motif (LysM) has been described in plants as a mediator of both symbiosis and immunity through direct binding of microbial glycans (reviewed in (*17*)). In the At5g62150 protein, it comprises residues 52-93 in (fig. S6). LysM recognizes bacterial glycan, chitin and presumably also other types of N-acetylglucosamine (GlcNAc) structures, which are common modifications of glycoproteins (*18*). However, At5g62150 itself does not contain predicted glycosylation sites and therefore is probably not a direct target of ARU (Chapter 3). We sequenced the *At5g62150* locus in the Lz-0, Ga-0, Fei-0, Sq-8 accessions but could not find any polymorphisms in the coding region (fig. S7). Thus, a loss-of-function in the high accessions Sq-8 and Fei-0 due to amino acid changes can be excluded. However, in general most of the variation in *A. thaliana* accessions is not caused by amino acid changes in proteins, but by expression level differences (*19*). Indeed, the accessions show several SNPs within 1000 bp up- and 800 bp downstream of the coding sequence, which could result in expression level differences (fig. S7).

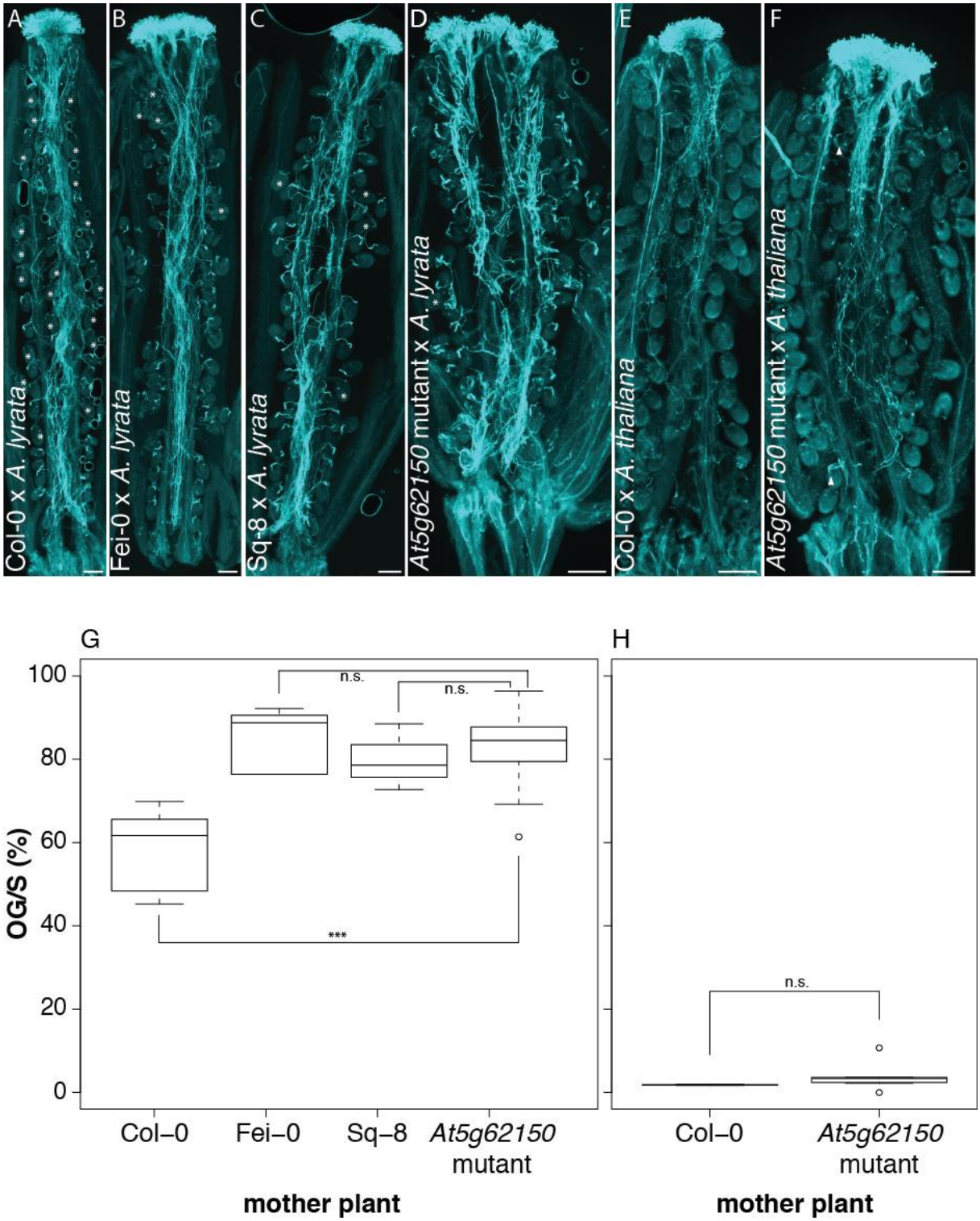


Figure 4.3: Mutants in *At5g62150* are impaired in interspecific, but not intraspecific, PT reception. (**A**) A Col-0 silique pollinated with *A. lyrata* pollen. Both ovules with overgrowth as well as normally fertilized ones (marked with asterisks) are visible. (**B and C**) Siliques of two accessions with very high OG/S pollinated with *A. lyrata* pollen. Only very few ovules are normally fertilized (asterisks). (B) Silique of Fei-0. (C) Silique of Sq-8. (**D**) A silique of the *At5g62150* mutant. Almost all ovules fail to recognize *A. lyrata* PTs (one normally fertilized ovule is marked with an asterisk). (**E**) A Col-0 silique pollinated with intraspecific *A. thaliana* pollen. All ovules receive a PT. (**F**) An *At5g62150* mutant silique pollinated with *A. thaliana* pollen. Almost all ovules are fertilized and only very few show overgrowth (arrowhead). (**G**) Ovules with interspecific PT overgrowth per silique (OG/S) in Col-0, Fei-0, Sq-8 and *At5g62150* mutants. **(continued pg. 140)**

In order to correlate the expression of *At5g62150* with the phenotypic difference between low and high accessions, we will assess mRNA levels using very sensitive digital droplet PCR (ddPCR), a very sensitive method that allows the detection of subtle differences in gene expression that could potentially be restricted to the synergid cells (*20*).

No signatures of positive selection are detectable in *At5g62150*

Many reproductive proteins involved in the direct recognition of male and female patterns are subject to rapid evolution and thus display signatures of positive selection (*21*). However, the amino acid sequence of At5g62150 is highly conserved between different Brassicaceae and we could not find such signatures our gene (using *At1g62150*, as well as its homologs in *A. lyrata, B. rapa, C. rubella, T. halophila*; fig. S6). Instead, we found an overall ratio of $d_N/d_S < 1$ (d_N/d_S = nonsynonymous substitutions per non-synonymous site/number of synonymous substitutions per synonymous site), indicative of purifying or stabilizing selection (*22*, *23*). In addition, we analyzed a 1852 bp fragment containing the coding sequence of *At5g62150* as well as surrounding regions (1001 bp up- and 542 bp downstream) in 84 *A. thaliana* accessions in intrapopulation tests for positive selection (Tajima's D and Fay and Wu's H, (*24*, *25*)). Although Tajima's D was negative and significantly different from the neutral null model (D=-1.76, p=0.014), the value does not lie within the 5% significance interval of D determined for 96 *A. thaliana* accessions which include all 84 used in this study (cut-off value: -2.08, (*26*)) and thus cannot be considered as an indicator of positive selection. Probably, the negative value of D is caused by demographic factors acting on the whole genome rahter than by positive selection on a single locus (*26*). Fay and Wu's H, another estimator of positive selection, was -1.10, however this value does not deviate from the neutral selection model (p=0.10). Taken together, these results suggest that *At5g62150* is not evolving under positive selection.

Figure Legend 4.3 (continued)
Mutants show significantly higher OG/S than Col-0 wild type (***p<0.0001, two-tailed t-test), but phenocopy Fei-0 and Sq-8 (difference is not significant, n.s.; p=0.5246 for the comparison of the mutant with Fei-0 and p=0.4770 for Sq-8). (**H**) Col-0 and *At5g62150* hand-pollinated with intraspecific *A. thaliana* Col-0 pollen. No significant difference was observed (p=0.4473). Scale Bars: 200 µm.

Discussion

Post-pollination, prezygotic hybridization barriers are an important mechanism for a plant species to prevent the formation of potentially sterile or unviable hybrid offspring. These barriers can act during each step of the fertilization process, namely during pollen germination, PT growth and guidance, as well as during PT reception and gamete fusion. All these processes strongly rely on cell-cell communication between the male gametophyte and the female tissues and are potential targets for species-recognition (*27*). In closely related species such as *A. thaliana* and *A. lyrata*, the species-recognition occurs during PT reception. If the *A. thaliana* embryo sacs do not recognize and actively reject *A. lyrata* PTs, the PTs do not release their sperm cells but instead continue growing inside the female gametophyte (PT overgrowth), leaving the female gametes unfertilized. The ability of *A. thaliana* embryo sacs to receive interspecific PTs is accession-dependent (Chapter 3). Conducting BSA with three F_2 populations originating from three different combinations of parental accessions, we were able to define a QTL spanning around 150 kb on chromosome 5. Within this region we identified *At5g62150*, coding for a small peptide with a LysM domain. Mutants in *At5g62150* have significantly higher OG/S in interspecific crosses than Col-0 wild type, thus phenocopying the high accessions Fei-0 and Sq-8. In plants, the LysM domain has mainly been implicated in regulating plant response to microorganisms (bacteria and fungi) during pathogen attack and symbiosis (reviewed in (*17*)). The characterized LysM domain-containing proteins are transmembrane receptor kinases or membrane-anchored extracellular co-receptors that can directly bind fungal or bacterial cell wall components (chitin and peptidoglycan, respectively). Interestingly, a shared role in PT reception and pathogen response has recently been described for the Mildew Locus O (MLO) protein NORTIA and the receptor-like kinase FERONIA (*7*). Both fungal and bacterial cell walls consist to a large proportion of N-acetylglucosamine (GlcNAc), which in plants is part of the carbohydrate oligomers that decorate glycoproteins. Previously, we have identified *ARU*, a gene coding for an oligosaccharyltransferase subunit involved in protein N-glycosylation (Chapter 3). Taken together, these results strengthen the significance of species-specific glycosylation patterns in the formation of hybridization barriers.

Material and Methods

Plant material and growth conditions

A. thaliana accessions and *A. lyrata* were grown as previously described (Chapter 3). T-DNA insertion mutants were obtained from the Nottingham Arabidopsis Stock Center (NASC). See table S5 for details.

Crosses, Aniline Blue staining and phenotyping

Buds of *A. thaliana* plants were emasculated and 2 days later, the pistils were pollinated with *A. lyrata* (interspecific) or *A. thaliana* (intraspecific) pollen. Two days after pollination, the siliques were harvested and fixed for Aniline Blue staining. Aniline Blue staining was conducted as previously described and PT overgrowth was analyzed using a Leica DM6000 epifluorescence microscope (Chapter 3). For the OG/S analysis in the F_1 generation, three siliques of three individual plants were assessed after interspecific pollination. For the initial segregation analysis in the F_2 generations, a minimum of three pollinated siliques of 27-44 individuals were counted. For the subsequent screening of the F_2 populations derived from Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0, a total number of 555, 586 and 620 individuals (with 2-3 siliques pollinated with *A. lyrata* pollen per individual) were screened, respectively. For OG/S analysis in candidate mutants, T-DNA lines (table S5) were genotyped and homozygous mutants were pollinated with *A. lyrata* pollen (2-3 plants per line for the initial screening). However, we were not able to genotype some of the lines (SALK_012418, SALK_043276, SALK_040581, SAIL_58_C10, SALK_020304, SALK_047222, SAIL_890_F09 and SAIL_823_B04). In these cases, we analyzed eight plants per line. If they did not show a phenotype, the mutant line was considered not to be involved in interspecific PT reception.

Sample preparation and Next-Generation Sequencing

Of each F_2 individual displaying either very high or very low proportions of PT overgrowth per silique, we collected two cauline leaves of about 1 cm length in separate pools for low and high phenotypes for each F_2 population (resulting in a total of two pools each for Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0). For the Fei-0 x Ga-0 population, we pooled material of 103 low individuals and 111 high ones (total

screened: 555), for Sq-8 x Ga-0, we collected leaves of 115 low and 118 high individuals (total screened: 586) and for Fei-0 x Lz-0, we bulked leaves of 131 low and 125 high plants (total screened: 620; table S3). The leaves were ground in liquid nitrogen using mortar and pestle and 100 mg material was used for DNA extraction with the Quiagen DNeasy Plant Mini Kit. Several rounds of DNA extraction were necessary to obtain 1 µg of DNA, which was required for Illumina HiSeq sequencing. Each of the six genomic sequencing libraries (Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0; low and high bulk each) was aligned to the ecotype-specific reference genomes (see table S6 (11, 12)) of their parents using bowtie (version 1.0.1, no mismatches and only unique alignments, (28)). Alignment statistics are shown in table S7. After obtaining all SNPs between two parental ecotypes, the alignments were used to calculate the occurrence of each SNP in the population. SNPs and their frequencies for all six samples can be found under https://www.dropbox.com/sh/zytsy3bdzf8t43a/AABIOD8xlVVuKhc_roxwJtPwa.

Bulk-Segregant Analysis

QTL mapping using bulk segregant analyses based on next generation sequencing was performed using a modified version of the method proposed by (*13*). The standard G statistic was calculated for every SNP using the following formula:

$$G = 2\sum_{i=1}^{4} n_i \ln\left(\frac{n_i}{\breve{n}_i}\right)$$

where $\breve{n}_i$ are the standard expected counts for a 2 x 2 contingency table under the null hypothesis that there are no QTLs linked to the SNP (e.g., $\breve{n}_1 = (n_1+n_2)(n_1+n_3)/(n_1+n_2+n_3+n_4)$, where n_1 and n_2, and n_3 and n_4, are the counts of the alleles from parent 1 and parent 2 in the low and the high bulks, respectively). The stochastic nature of next-generation sequencing causes random variation in the number of reads per allele within each bulk and, consequently, substantial variation in G. To reduce this variation, G', a smoothed version of G, was calculated by averaging G values across neighboring SNPs using a sliding window with a fixed width of 5000 SNPs.

The significance threshold of G' was estimated using a non-parametric empirical approach originally proposed by (*13*). The observed distribution of G' is a mixture of the null distribution (regions not linked to QTLs) and several contaminating distributions (regions linked to QTLs). The null distribution of G' is expected to be right-skewed and

asymptotically follow a log-normal distribution, $\theta_{G'} \sim \ln N(\mu, \sigma^2)$. The parameters of the $\theta_{G'}$ null distribution can be empirically estimated using the parameters of the observed G' distribution ($X_{G'}$) with potential QTL-harboring regions removed. The G' trimmed dataset (X_T) was constructed with SNPs that satisfy:

$$w_i - \text{Median}(W_{G'}) \leq g(N, \alpha_N)\text{MAD}_l(W_{G'})$$

where $W_{G'} = \ln[X_{G'}]$ (logarithm of observed data), $g(N, \alpha_N)$ defines the limits of the outlier regions and was taken to be 5.2 (equivalent to observations with p-values <0.001 for normally distributed data), and MAD_l is the left median absolute deviation of $W_{G'}$. The median and mode of X_T were then used to estimate the $\theta_{G'}$ null distribution parameters $\mu = \ln[\text{Median}(X_T)]$ and $\sigma^2 = \mu - \ln[\text{Mode}(X_T)]$. The estimated $\theta_{G'}$ null distribution was used to calculate the threshold for G' at a 1% false discovery rate.

GWAS analysis

Multi-locus mixed model (MLMM) analysis was conducted using the method developed by (*15*) using mean values of interspecific PT overgrowth for 86 accessions (Chapter 3).

LysM domain identification and glycosylation site prediction

The LysM domain in At5g62150 could not be identified using the Pfam website (http://pfam.xfam.org/), however it could be found using other motif search tools and the Pfam database (http://www.genome.jp/tools/motif/, http://hits.isb-sib.ch/cgi-bin/PFSCAN). Glycosylation site prediction was conducted with the NetGlyc 1.0 server using standard settings (http://www.cbs.dtu.dk/services/NetNGlyc/).

Alignments and sequencing of At5g62150 in A. thaliana accessions and Brassicaceae

We amplified the coding region of *At5g62150* in Lz-0, Ga-0, Fei-0 and Sq-8 including 1000 bp of upstream and 810 bp of downstream sequence using the gene-specific primers 5' – AAGGTAGTTGTAACGTTGG – 3' and 5' – ATCCCTACTTACCCATATC – 3'. PCR products were cloned into the pJET 1.2 cloning vector using the CloneJet PCR Cloning Kit (Thermo Scientific) and sequenced. Alignments were conducted using ClustalX. Homologous *At1g62150* sequences from the Brassicaceae *Arabidopsis lyrata, Capsella*

rubella, Brassica rapa and *Thellungiella halophila* were downloaded from www.phytozome.net.

Positive Selection analysis

d_N/d_S was calculated using MEGA 5.2.2. Intrapopulation selection tests were conducted with http://wwwabi.snv.jussieu.fr/achaz/neutralitytest.html. For a list of the 84 accessions that were used, see table S8. *A. lyrata* was used as outgroup and p-values were determined by conducting 100000 simulations. 5% significance threshold for Tajima's D was calculated with data for 96 accessions from (*26*). The 84 accessions that were used are all included in the 96 accession data set, however for the exact comparison of our calculation with the genome-wide distribution of D obtained in (*26*), the analysis has to be repeated with the full set of 96 accessions.

References

1. A. Widmer, C. Lexer, S. Cozzolino, Evolution of reproductive isolation in plants, *Heredity* **102**, 31–38 (2009).

2. S. A. Kessler, U. Grossniklaus, She's the boss: signaling in pollen tube reception, *Current opinion in plant biology* **14**, 622–627 (2011).

3. A. R. Leydon *et al.*, Three MYB transcription factors control pollen tube differentiation required for sperm release, *Current biology* **23**, 1209–1214 (2013).

4. N. Huck, J. M. Moore, M. Federer, U. Grossniklaus, The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception, *Development* **130**, 2149–2159 (2003).

5. A. Boisson-Dernier, S. Frietsch, T. H. Kim, M. B. Dizon, J. I. Schroeder, The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition, *Current biology : CB* **18**, 63–68 (2008).

6. A. Capron *et al.*, Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene, *The Plant cell* **20**, 3038–3049 (2008).

7. S. A. Kessler *et al.*, Conserved molecular components for pollen tube reception and fungal invasion, *Science* **330**, 968–971 (2010).

8. E. G. Williams, V. Kaul, J. L. Rouse, B. F. Palser, Overgrowth of Pollen Tubes in Embryo Sacs of *Rhododendron* Following Interspecific Pollinations, *Aust. J. Bot.* **34**, 413–423 (1986).

9. J. M. Escobar-Restrepo *et al.*, The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception, *Science* **317**, 656–660 (2007).

10. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, *Nature* **465**, 627–631 (2010).

11. J. Cao *et al.*, Whole-genome sequencing of multiple *Arabidopsis thaliana* populations, *Nature genetics* **43**, 956–963 (2011).

12. M. W. Horton *et al.*, Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel, *Nature genetics* **44**, 212–216 (2012).

13. P. M. Magwene, J. H. Willis, J. K. Kelly, The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing, *PLoS Comput. Biol.* **7** (2011).

14. Z. Yang *et al.*, Mapping of Quantitative Trait Loci Underlying Cold Tolerance in Rice Seedlings via High-Throughput Sequencing of Pooled Extremes, *PLoS ONE* **8**, e68433 (2013).

15. V. Segura *et al.*, An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, *Nature genetics* **44**, 825–830 (2012).

16. S. E. Wuest *et al.*, *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes, *Current biology* **20**, 506–512 (2010).

17. A. A. Gust, R. Willmann, Y. Desaki, H. M. Grabherr, T. Nürnberger, Plant LysM proteins: modules mediating symbiosis and immunity, *Trends Plant Sci.* **17**, 495–502 (2012).

18. M. Aebi, N-linked protein glycosylation in the ER, *Biochimica et biophysica acta* **1833**, 2430–2437 (2013).

19. X. Gan *et al.*, Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*, *Nature* **477**, 419–423 (2011).

20. B. J. Hindson *et al.*, High-throughput droplet digital PCR system for absolute quantitation of DNA copy number, *Anal. Chem.* **83**, 8604–8610 (2011).

21. W. J. Swanson, V. D. Vacquier, The rapid evolution of reproductive proteins, *Nature reviews. Genetics* **3**, 137–144 (2002).

22. M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Molecular Biology and Evolution* **3**, 418–426 (1986).

23. C. F. Mugal, J. B. W. Wolf, I. Kaj, Why time matters: codon evolution and the temporal dynamics of dN/dS, *Molecular Biology and Evolution* **31**, 212–231 (2014).

24. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism, *Genetics* **123**, 585–595 (1989).

25. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection, *Genetics* **155**, 1405–1413 (2000).

26. M. Nordborg *et al.*, The pattern of polymorphism in *Arabidopsis thaliana*, *PLoS biology* **3**, e196 (2005).

27. R. Swanson, A. F. Edlund, D. Preuss, Species Specificity In Pollen-Pistil Interactions, *Annu. Rev. Genet.* **38**, 793–818 (2004).

28. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.* **10**, R25 (2009).

Supplementary Figures and Tables

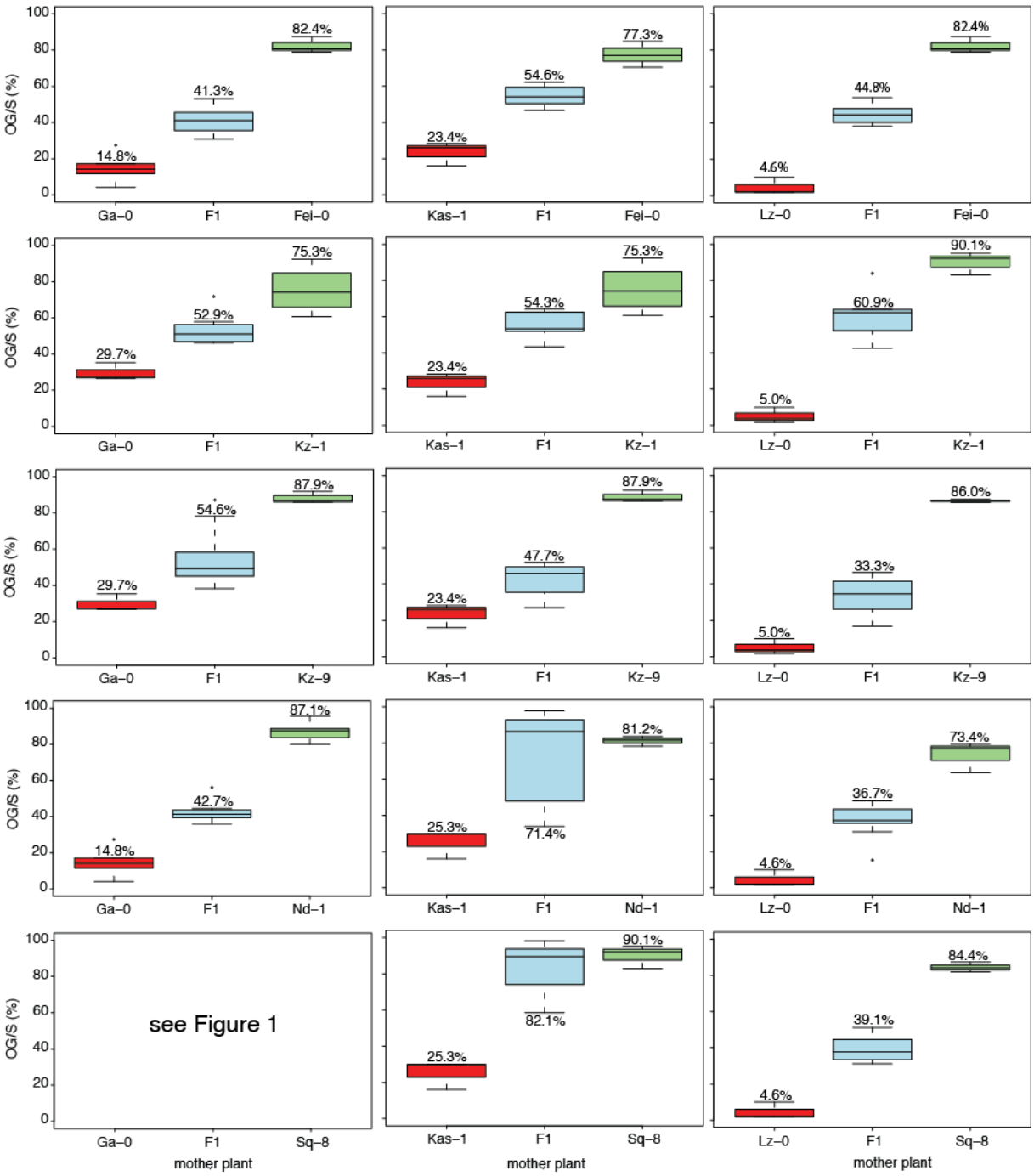
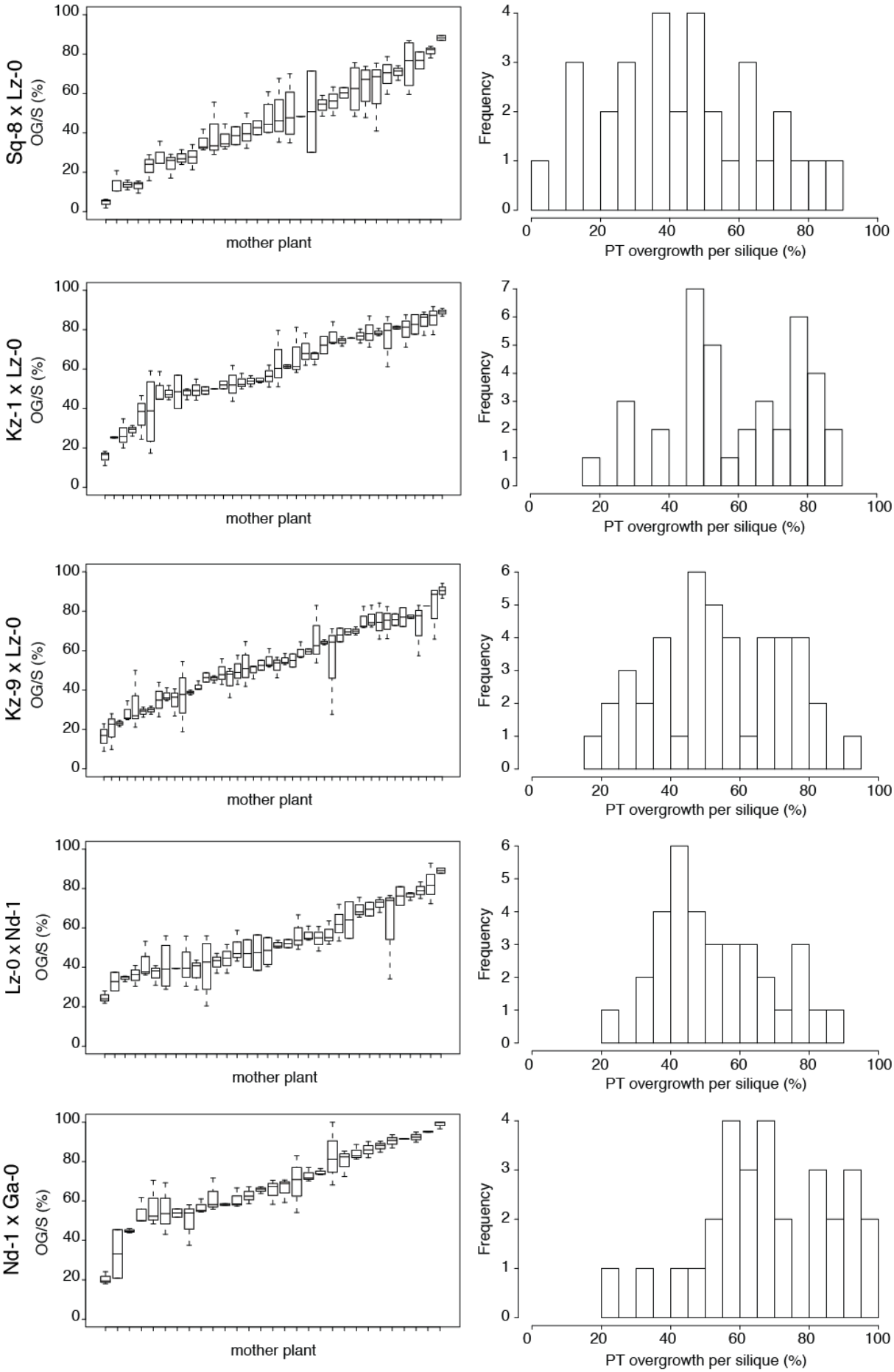


Figure S1: F_1 accession hybrids and their parental lines in crosses with *A. lyrata*. OG/S for F_1 is marked in blue, and the respective low and high parents are color-coded in red and green in the same panel. Values near the box plots denote mean values of OG/S. OG/S values for several parental accessions have been plotted more than once in different panels (same OG/S means). These values originate from the same experiment and were plotted near the respective F_1 and the other parent to facilitate understanding.


Sq-8 x Lz-0
OG/S (%)
mother plant
Frequency
PT overgrowth per silique (%)
Kz-1 x Lz-0
OG/S (%)
mother plant
Frequency
PT overgrowth per silique (%)
Kz-9 x Lz-0
OG/S (%)
mother plant
Frequency
PT overgrowth per silique (%)
Lz-0 x Nd-1
OG/S (%)
mother plant
Frequency
PT overgrowth per silique (%)
Nd-1 x Ga-0
OG/S (%)
mother plant
Frequency
PT overgrowth per silique (%)

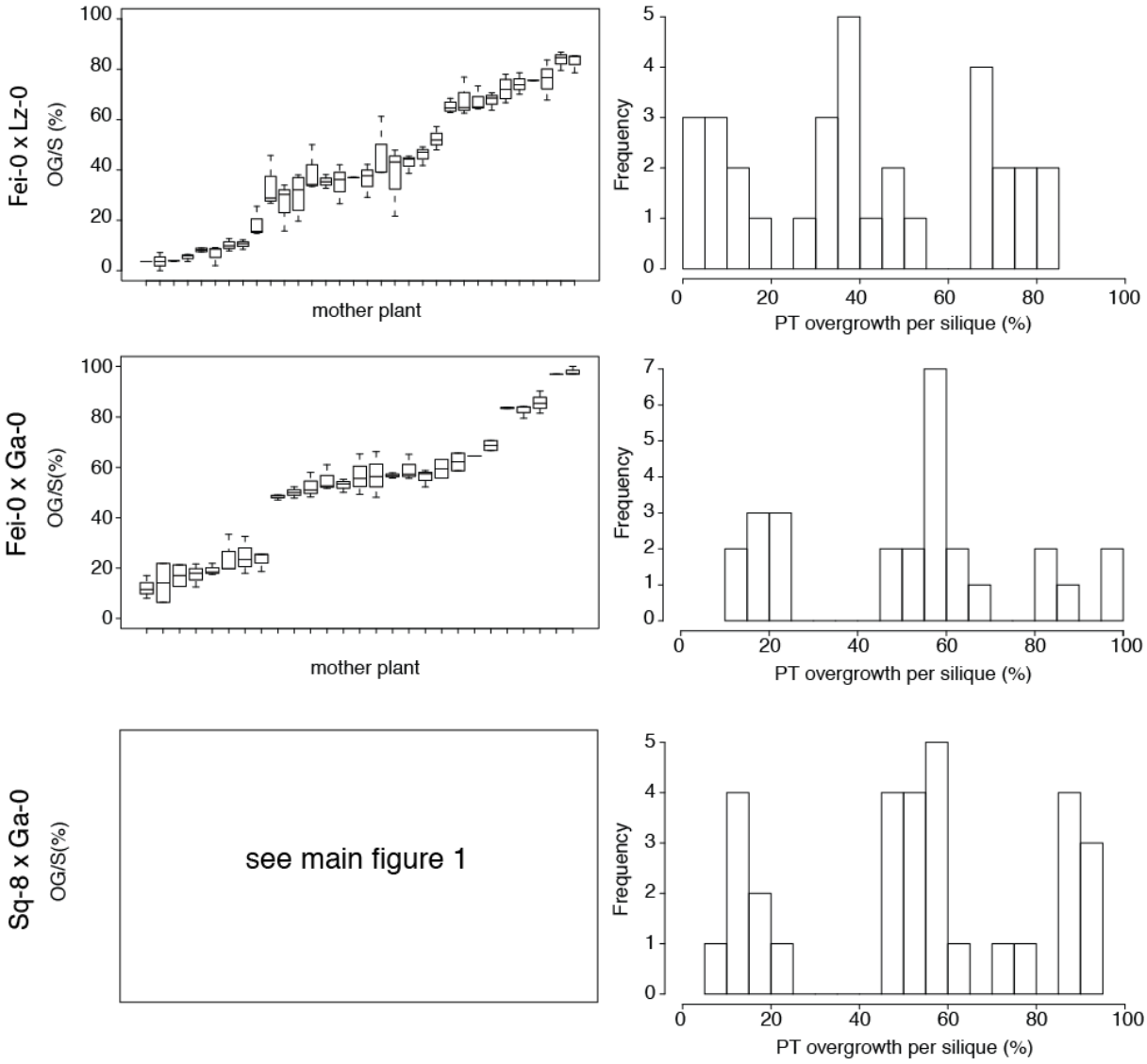


Figure S2: Segregation analysis of F_2 populations derived from eight combinations of parental accessions. OG/S values were collected for 27-44 F_2 individuals per population. The left panel depicts the raw data for each mother plant, and the right panel is a histogram for OG/S across a population. The F_2 populations segregate into three distinct groups (low, intermediate and high OG/S) for Fei-0 x Lz-0, Fei-0 x Ga-0 and Sq-8 x Ga-0, but show a gradual distribution for all other parental crosses.

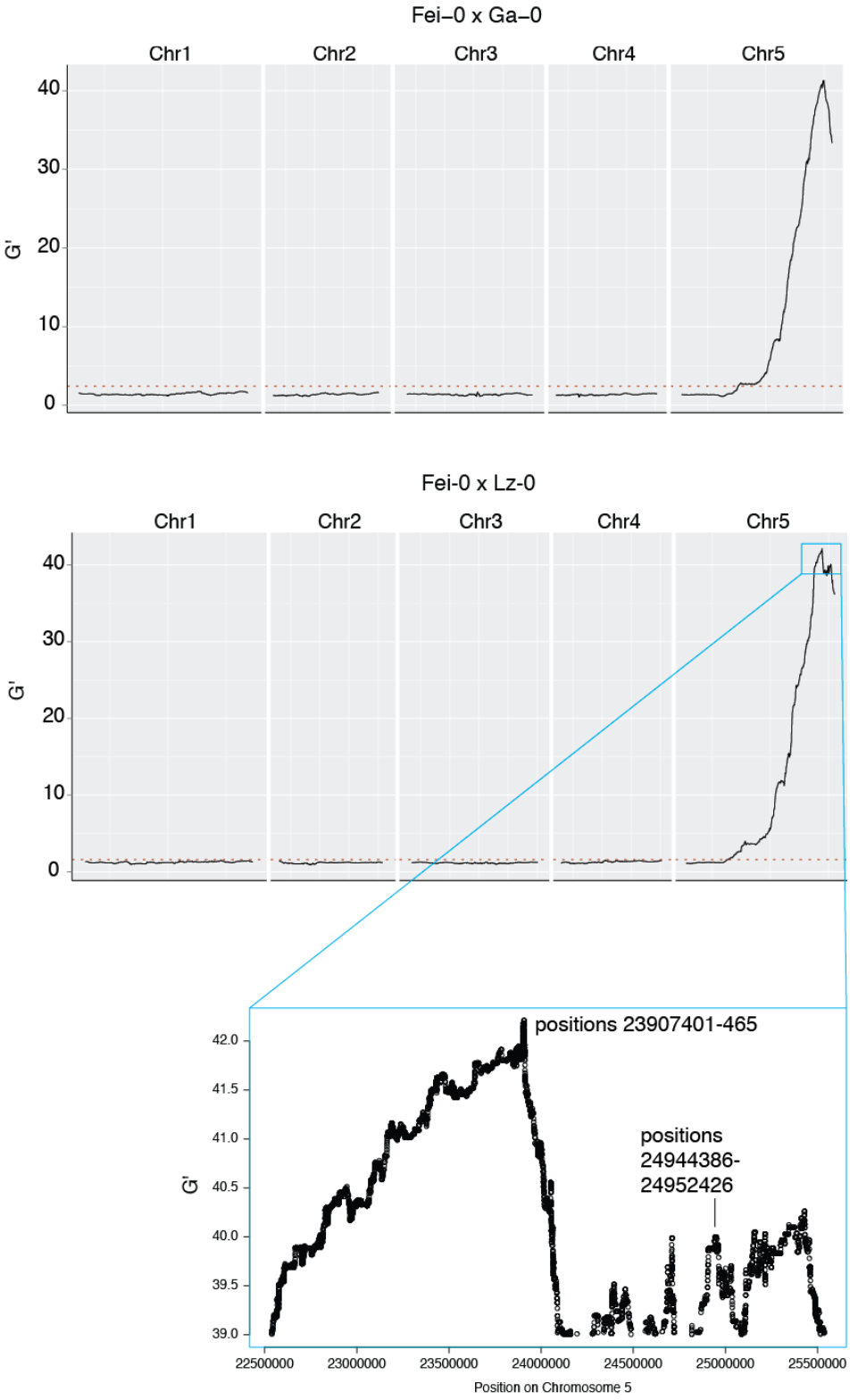


Figure S3: BSA peaks for Fei-0 x Ga-0 and Fei-0 x Lz-0. Both accession combinations point to the same region on chromosome 5 as identified for Sq-8 x Ga-0. The Fei-0 x Lz-0 main peak is about 1.1 Mbp away from the peaks of the other two mapping populations, but several small peaks are visible (blue box). One of these smaller peaks (maximal G' values at positions 24944386-24952426) is very close to the position of the main peaks of Fei-0 x Ga-0 and Sq-8 x Ga-0.

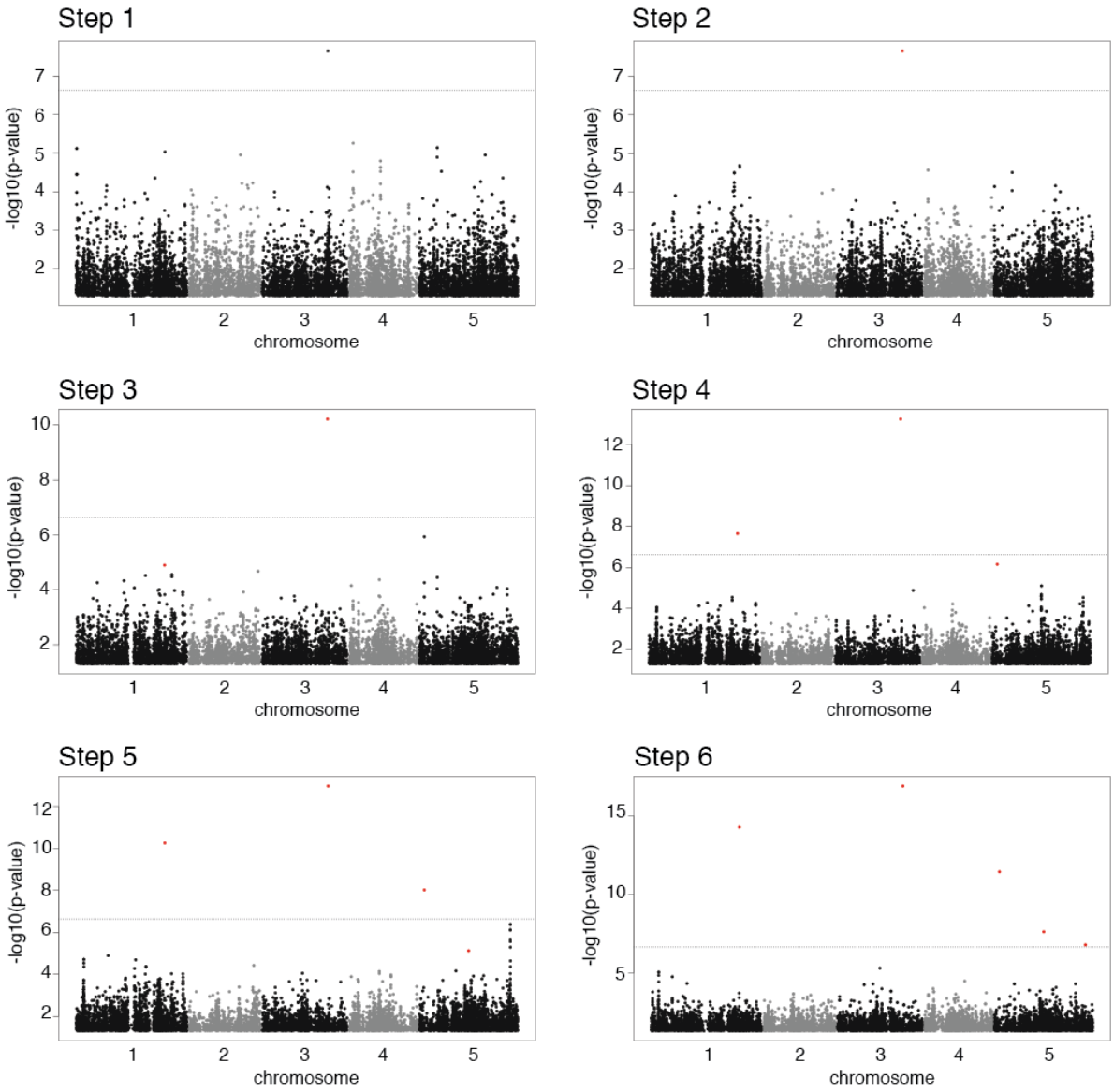


Figure S4: GWAS multi-locus mixed model (MLMM) analysis on OG/S values of 86 accessions (Chapter 3). From Step 2 on, the SNP with the lowest p-value from the previous calculation (step) was used as a co-factor in the model (marked in red). In step 5, a prominent peak on chromosome 5 is visible. This peak points to the same region obtained in the BSA analysis. Dashed line: 5% Bonferroni-corrected p-value.

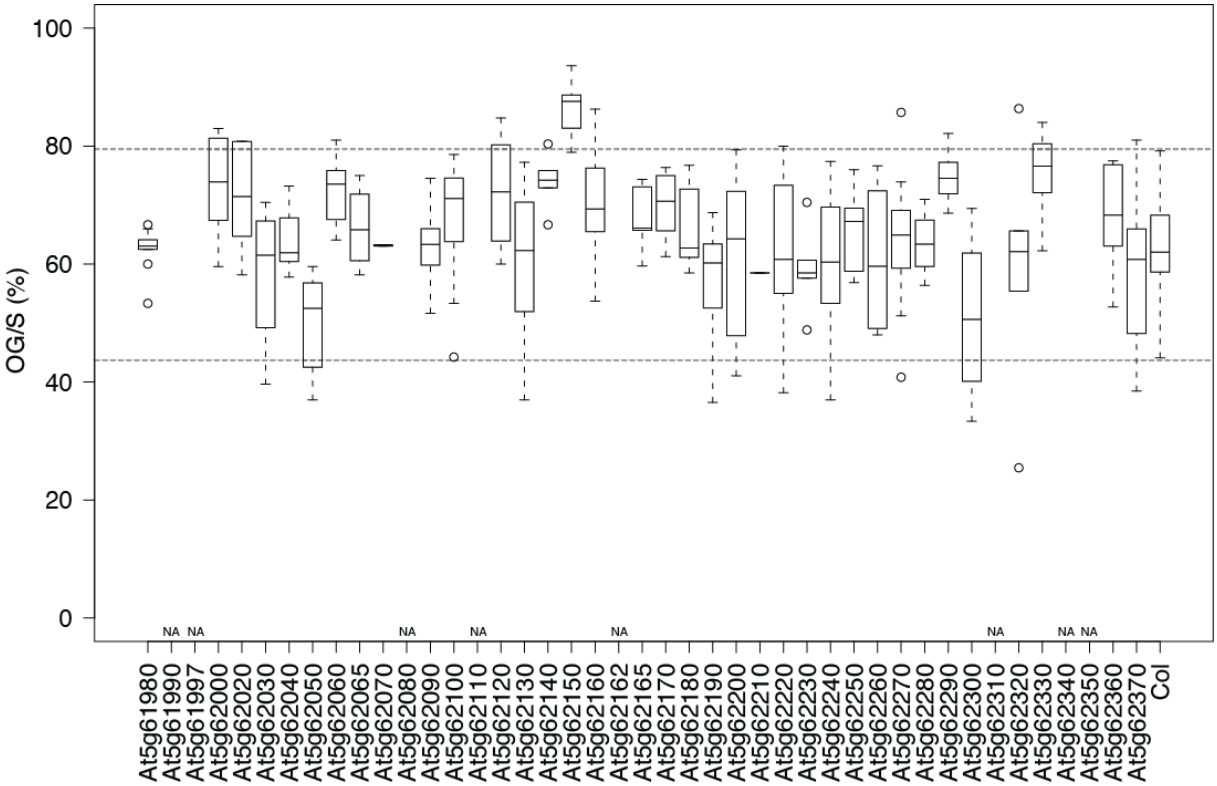


Figure S5: Mutants in candidate genes pollinated with *A. lyrata*. For each gene in the 150 kb candidate region (if available), one mutant line was used as mother plant in interspecific crosses. The dashed line corresponds to the variation in OG/S determined by the Col-0 control. Only the mutant in *At5g62150* clearly deviates from this distribution. NA: no T-DNA line available.

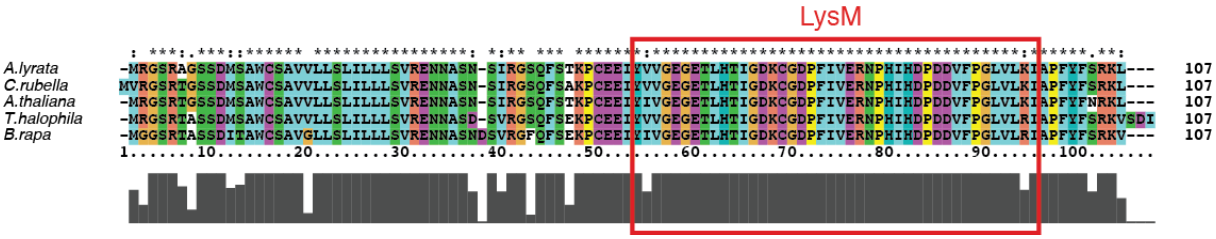
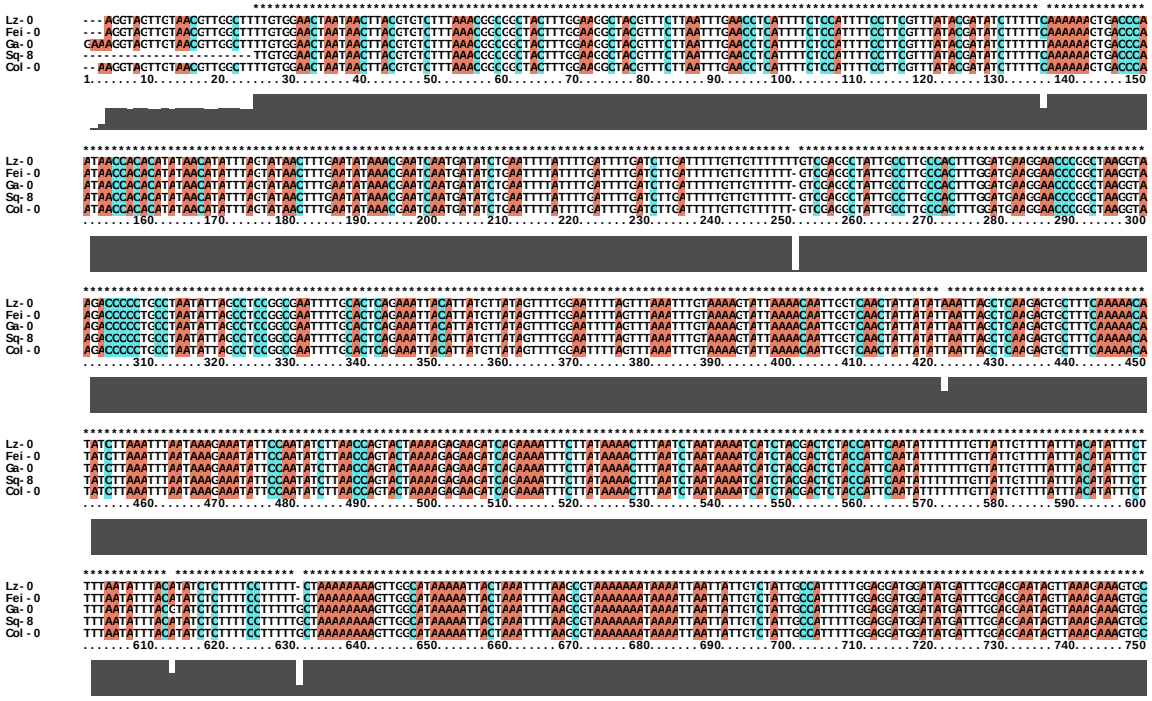
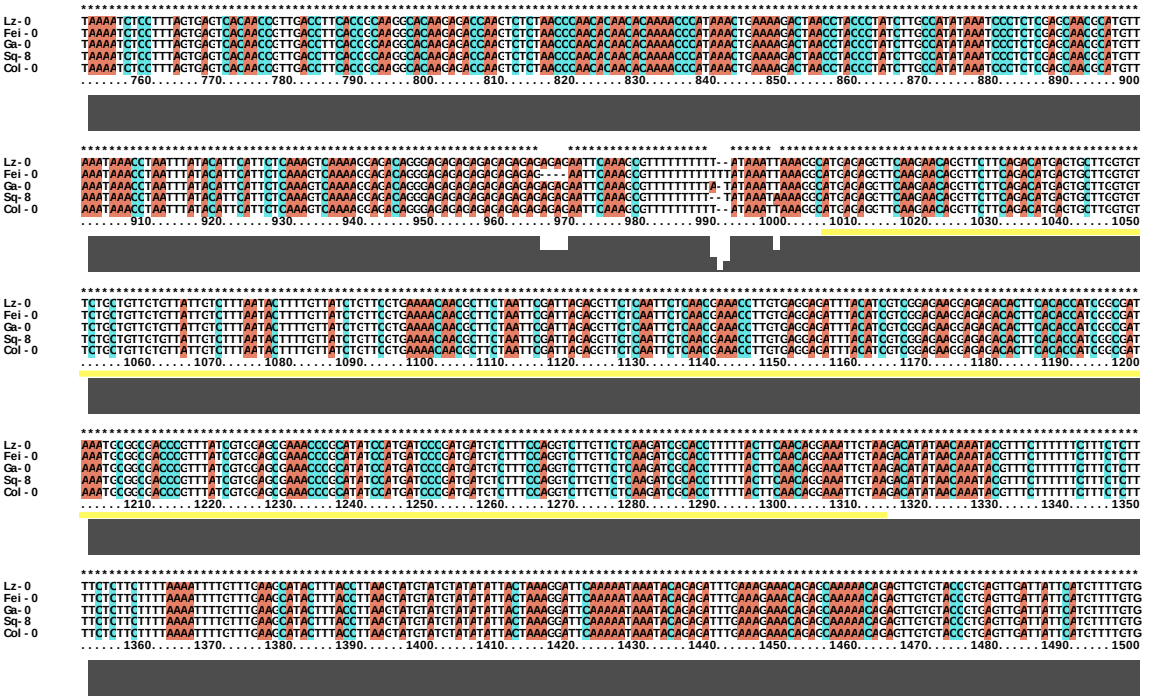


Figure S6: Alignment of *At5g62150* and its homologs in related Brassicaceae. The conserved LysM domain is highlighted in red.

Page 1 of 3



Page 2 of 3



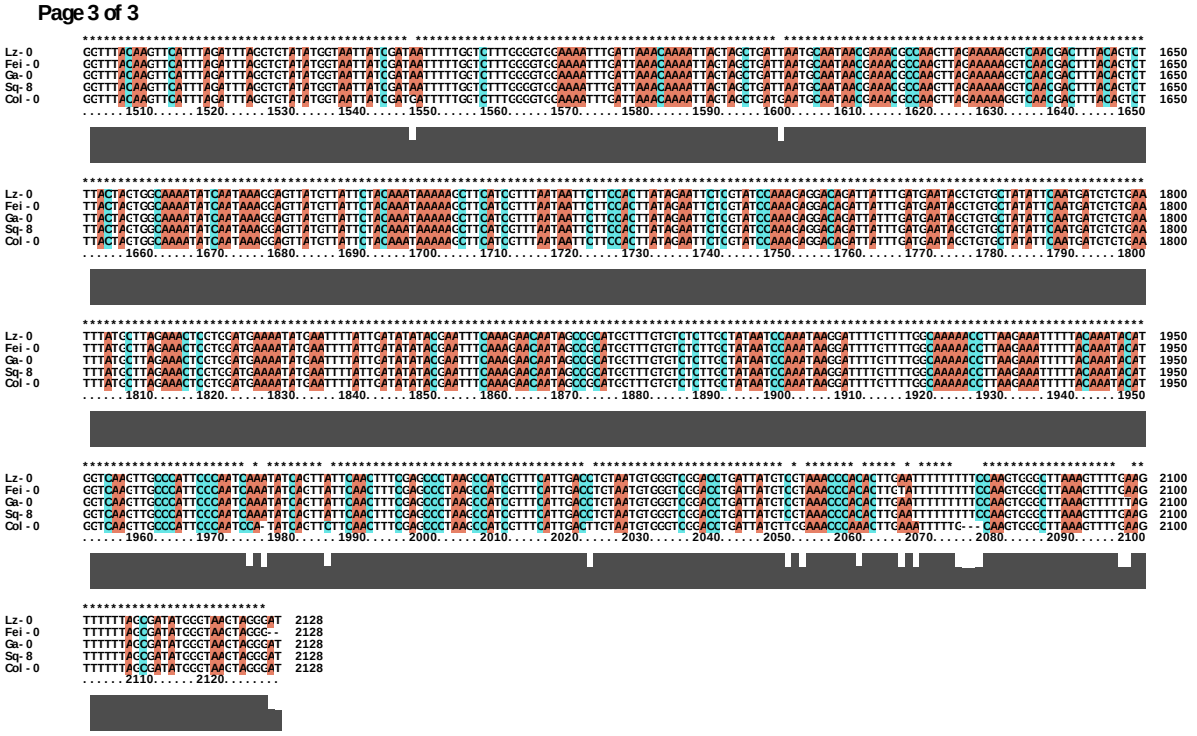


Figure S7: Alignment of *At5g62150* and surrounding genomic regions in Lz-0 and Ga-0 (low OG/S), Fei-0 and Sq-8 (high OG/S) as well as Col-0 (reference). The coding region of *At5g62150* is marked with a yellow bar. No SNPs were found in the coding region, and the SNPs up- and downstream do not show an obvious pattern that could be correlated to the OG/S phenotype.

Table S1: Combinations of accessions used for the generation of segregating F_2 populations and their phenotypes (OG/S) and segregation ratios (low:intermediate:high) in F_1 and F_2, respectively.

Female parent	Male parent	F1 phenotype	F2 segregation
Fei-0	Ga-0	intermediate	1:2:1
Fei-0	Kas-1	intermediate	n.d.
Fei-0	Lz-0	intermediate	1:2:1
Kz-1	Ga-0	intermediate	n.d.
Kz-1	Kas-1	intermediate	n.d.
Kz-1	Lz-0	intermediate	quantitative
Kz-9	Ga-0	intermediate	n.d.
Kz-9	Kas-1	intermediate	n.d.
Kz-9	Lz-0	intermediate	quantitative
Sq-8	Ga-0	intermediate	1:2:1
Sq-8	Kas-1	high	n.d.
Sq-8	Lz-0	intermediate	quantitative
Nd-1	Ga-0	intermediate	quantitative
Kas-1	Nd-1	intermediate (-high	n.d.
Lz-0	Nd-1	intermediate	quantitative

Table S2: OG/S segregation in selected F_2 populations. (A) Numbers of F_2 individuals corresponding to the low, the intermediate or the high OG/S group. Individuals were assigned to low and high groups that correspond to 95% confidence intervals for the respective parental accessions that are shown in **(B)**. 95% Confidence intervals were calculated in R using a t-distribution with OG/S data for Fei-0, Ga-0, Lz-0 and Sq-8 obtained in our initial experiment (Chapter 3).

A

F2 Populations and their Assignment to Low, High and Intermediate Phenotype Groups						
F2 population	**Low Range (%)**	**# of Individuals**	**Intermediate Range (%)**	**# of Individuals**	**High Range (%)**	**# of Individuals**
Sq-8 x Ga-0	5.93 – 21.75	8	45.60 – 76.44	16	86.42 – 90.89	7
Fei-0 x Ga-0	12.13 – 24.60	8	48.24 – 68.63	14	82.47 – 98.04	5
Fei-0 x Lz-0	3.57 – 10.45	8	18.64 – 68.00	18	72.14 – 83.65	6

B

95% Confidence Interval OG/S (%) *	
Lz-0	7.58 – 12.47
Ga-0	12.37 – 24.12
Fei-0	71.75 – 84.53
Sq-8	77.50 – 89.39

*** calculated from our previous results (Müller et al., in preparation).**

The confidence intervals in (B) were used to group the F2 individuals according to their phenotype into a low and a high group (A).

Table S3: Phenotyping for Next Generation Sequencing. Three F_2 populations were screened and tissue of individuals with low and high OG/S was collected for DNA extraction in a low and a high pool, respectively. The table shows both the observed and the expected values for a 1:2:1 segregation (low:intermediate:high). In all cases, the expected values for low and high are higher than the actual ones. This is because we only picked very clear phenotypes for next-generation sequencing and borderline OG/S were left out in order to avoid sampling mistakes.

Original(Cross	Total(Number(of(Plants	Low(Pool	Intermediate	High(Pool
Fei<0(x(Lz<0	**620**	**131**	**364**	**125**
expected'for'1:2:1	620	155	310	155
Fei<0(x(Ga<0	**555**	**103**	**341**	**111**
expected'for'1:2:1	555	138.75	277.5	138.75
Sq<8(x(Ga<0	**586**	**115**	**353**	**118**
expected'for'1:2:1	586	146.5	293	146.5

Table S4: Results of GWAS MLMM. The top 25 associated SNPs with their PERL identifiers, their location in the genome and their p-value are shown for each step.

step1

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	2,24E-08	*AT3G48600*
2	4	979996	PERL0667390	5,53E-06	*AT4G02230*
3	5	4924090	PERL0891205	7,24E-06	outside gene
4	1	119003	PERL0000734	7,60E-06	outside gene
5	1	24325910	PERL0219430	9,29E-06	outside gene
6	2	14123516	PERL0382632	1,11E-05	*AT2G33330*
7	5	18084064	PERL1043650	1,12E-05	*AT5G44800*
8	5	4899000	PERL0891072	1,28E-05	*AT5G15120*
9	4	8499071	PERL0759382	1,59E-05	*AT4G14790*
10	4	8497876	PERL0759366	2,34E-05	*AT4G14790*
11	4	8501077	PERL0759396	2,34E-05	*AT4G14800*
12	4	8502544	PERL0759416	2,34E-05	*AT4G14805*
13	5	6068770	PERL0901506	2,97E-05	*AT5G18330*
14	4	8507406	PERL0759533	2,98E-05	*AT4G14819*
15	4	972178	PERL0667236	3,05E-05	*AT4G02195*
16	1	133186	PERL0000813	3,55E-05	*AT1G01340*
17	1	133626	PERL0000822	3,55E-05	*AT1G01340*
18	1	133701	PERL0000823	3,55E-05	*AT1G01340*
19	1	134063	PERL0000832	3,55E-05	*AT1G01340*
20	5	22934169	PERL1095958	4,40E-05	*AT5G56660*
21	1	21605799	PERL0183896	4,43E-05	*AT1G58260*
22	5	16827147	PERL1022663	5,47E-05	*AT5G42090*
23	2	17494583	PERL0401639	5,92E-05	outside gene
24	2	14516942	PERL0384681	6,06E-05	*AT2G34400*
25	4	8504510	PERL0759481	6,26E-05	*AT4G14810*

step2

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	2,19E-08	*AT3G48600*
2	1	24277120	PERL0218693	2,06E-05	*AT1G65350*
3	1	24325910	PERL0219430	2,30E-05	outside gene
4	4	979996	PERL0667390	2,72E-05	*AT4G02230*
5	5	4924090	PERL0891205	3,13E-05	*AT5G56330*
6	1	22816028	PERL0198560	3,18E-05	outside gene
7	1	22842905	PERL0198961	3,21E-05	outside gene
8	1	22798368	PERL0198356	5,74E-05	*AT1G61740*
9	5	16767272	PERL1021646	6,95E-05	outside gene
10	5	58898	PERL0853591	7,25E-05	*AT5G01170*
11	1	22814316	PERL0198538	7,41E-05	outside gene
12	2	19055605	PERL0412330	8,76E-05	*AT2G46420*
13	1	22841729	PERL0198924	9,16E-05	*AT1G61830*
14	5	4899000	PERL0891072	9,25E-05	*AT5G15120*
15	1	22509374	PERL0194320	9,67E-05	*AT1G61100*

16	1	22509563	PERL0194321	9,67E-05	*AT1G61100*
17	5	18084064	PERL1043650	9,94E-05	*AT5G44800*
18	2	15995411	PERL0392478	0,000107457	outside gene
19	1	6631599	PERL0048466	0,000125048	*AT1G19220*
20	1	22795841	PERL0198277	0,000128594	*AT1G19220*
21	4	18450831	PERL0852268	0,000140745	*AT4G39770*
22	1	27020437	PERL0245490	0,000141473	*AT1G71830*
23	4	972178	PERL0667236	0,000154364	*AT4G02195*
24	5	16714592	PERL1020649	0,000164388	outside gene

step3

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	6,06E-11	*AT3G48600*
2	5	1369413	PERL0866477	1,18E-06	*AT5G04740*
3	1	24277120	PERL0218693	1,27E-05	*AT1G65350*
4	2	19055605	PERL0412330	2,11E-05	*AT2G46420*
5	1	26296573	PERL0239571	2,78E-05	outside gene
6	1	19049015	PERL0162844	3,02E-05	*AT1G51380*
7	1	26294785	PERL0239549	3,34E-05	*AT1G69840*
8	5	4924090	PERL0891205	3,54E-05	outside gene
9	4	8209018	PERL0755210	4,29E-05	*AT4G14250*
10	1	13117521	PERL0112360	4,64E-05	outside gene
11	1	5833476	PERL0042361	5,49E-05	*AT1G17060*
12	5	1370695	PERL0866496	5,49E-05	*AT5G04740*
13	4	505633	PERL0661920	7,05E-05	*AT4G01190*
14	5	21299239	PERL1082073	8,18E-05	*AT5G52490*
15	1	15974746	PERL0130427	8,45E-05	*AT1G42540*
16	5	4899000	PERL0891072	9,10E-05	*AT5G15120*
17	5	24129748	PERL1105590	9,11E-05	*AT5G59920*
18	1	22842905	PERL0198961	9,83E-05	outside gene
19	1	22521533	PERL0194448	0,00010382	outside gene
20	1	27020437	PERL0245490	0,000105503	*AT1G71830*
21	2	15019298	PERL0387191	0,000120757	*AT2G35720*
22	1	29282670	PERL0260616	0,000123588	*AT1G77860*
23	1	29282732	PERL0260617	0,000123588	*AT1G77860*
24	1	29284306	PERL0260630	0,000123588	*AT1G77860*

step4

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	5,82E-14	*AT3G48600*
2	1	24277120	PERL0218693	2,23E-08	*AT1G65350*
3	5	1369413	PERL0866477	7,01E-07	*AT5G04740*
4	5	13534683	PERL0977245	7,83E-06	*AT5G35332*
5	3	21459702	PERL0644748	1,30E-05	outside gene
6	5	13490741	PERL0976715	2,06E-05	*AT5G35230*
7	5	13478276	PERL0976642	2,51E-05	*AT5G35210*
8	5	13478953	PERL0976649	2,51E-05	*AT5G35210*
9	1	22841729	PERL0198924	2,82E-05	*AT1G61830*
10	5	24994298	PERL1111187	2,89E-05	outside gene
11	5	13496330	PERL0976790	3,41E-05	*AT5G35240*
12	1	22842107	PERL0198933	3,90E-05	outside gene

13	1	22842905	PERL0198961	3,92E-05	outside gene
14	5	24996125	PERL1111205	4,74E-05	outside gene
15	5	25000607	PERL1111214	4,74E-05	*AT5G62230*
16	1	15974746	PERL0130427	5,24E-05	*AT1G42540*
17	1	29284809	PERL0260638	5,54E-05	outside gene
18	5	24125192	PERL1105517	5,93E-05	*AT5G59900*
19	4	8288507	PERL0756357	5,99E-05	*AT4G14385*
20	1	29282670	PERL0260616	6,89E-05	*AT1G77860*
21	1	29282732	PERL0260617	6,89E-05	*AT1G77860*
22	1	29284306	PERL0260630	6,89E-05	*AT1G77860*
23	1	29285278	PERL0260647	6,89E-05	*AT1G77870*
24	5	10027931	PERL0946958	7,34E-05	*AT5G28020*
25	1	13692596	PERL0120331	7,39E-05	*AT1G36360*

step5

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	1,07E-13	*AT3G48600*
2	1	24277120	PERL0218693	5,51E-11	*AT1G65350*
3	5	1369413	PERL0866477	9,53E-09	*AT5G04740*
4	5	24996125	PERL1111205	4,21E-07	ouside gene
5	5	25000607	PERL1111214	4,21E-07	*AT5G62230*
6	5	24983562	PERL1111097	7,69E-07	*AT5G62190*
7	5	24984653	PERL1111101	7,69E-07	*AT5G62200*
8	5	24984851	PERL1111104	7,69E-07	*AT5G62200*
9	5	24979764	PERL1111068	2,12E-06	*AT5G62180*
10	5	24994298	PERL1111187	2,75E-06	ouside gene
11	5	25004234	PERL1111229	5,12E-06	*AT5G62240*
12	5	13534683	PERL0977245	7,50E-06	*AT5G35332*
13	1	8729332	PERL0067718	1,28E-05	*AT1G24640*
14	1	2092016	PERL0013520	1,96E-05	*AT1G06810*
15	1	2093900	PERL0013532	1,96E-05	*AT1G06820*
16	1	16279927	PERL0133665	2,04E-05	ouside gene
17	5	25021366	PERL1111289	2,28E-05	*AT5G62300*
18	1	2073665	PERL0013415	2,92E-05	*AT1G06750*
19	2	17730979	PERL0403583	3,78E-05	*AT2G42580*
20	1	19124602	PERL0163397	4,20E-05	*AT1G51570*
21	5	25014054	PERL1111266	4,35E-05	*AT5G62270*
22	1	2095951	PERL0013541	4,40E-05	*AT1G06820*
23	1	19133268	PERL0163430	4,41E-05	*AT1G51600*
24	1	19139420	PERL0163451	4,41E-05	*AT1G51610*
25	1	15974746	PERL0130427	5,25E-05	*AT1G42540*

step6

Rank	Chr	Position	SNPname	p-value	Gene
1	3	18010190	PERL0617038	1,27E-17	*AT3G48600*
2	1	24277120	PERL0218693	5,26E-15	*AT1G65350*
3	5	1369413	PERL0866477	3,67E-12	*AT5G04740*
4	5	13534683	PERL0977245	2,42E-08	*AT5G35332*
5	5	24996125	PERL1111205	1,65E-07	outside gene
6	3	11700017	PERL0529369	5,01E-06	*AT3G29792*
7	1	2092016	PERL0013520	9,11E-06	*AT1G06810*
8	1	2093900	PERL0013532	9,11E-06	*AT1G06820*
9	1	2073665	PERL0013415	1,44E-05	*AT1G06750*
10	1	5836785	PERL0042378	1,77E-05	outside gene
11	4	10971668	PERL0795782	3,33E-05	outside gene
12	1	2095951	PERL0013541	3,74E-05	*AT1G06820*
13	1	9871467	PERL0076124	4,71E-05	*AT1G28240*
14	5	22280344	PERL1090766	5,05E-05	outside gene
15	5	5451698	PERL0895286	5,11E-05	*AT5G16620*
16	3	9784282	PERL0500530	5,18E-05	outside gene
17	3	7370968	PERL0473400	5,70E-05	*AT3G21030*
18	5	16929094	PERL1023967	8,43E-05	*AT5G42340*
19	4	2294870	PERL0683166	9,99E-05	*AT4G04580*
20	3	12237278	PERL0537478	0,000102171	outside gene
21	5	16063229	PERL1012591	0,000116962	*AT5G40160*
22	5	16065219	PERL1012608	0,000116962	*AT5G40170*
23	5	9544872	PERL0940649	0,000127053	outside gene
24	1	2076796	PERL0013424	0,000143675	*AT1G06760*
25	5	12514868	PERL0968925	0,000144196	*AT5G33253*

Table S5: The 150 kb candidate region described by the BSA peaks and the GWAS MLMM peak. This table shows the gene IDs, their predicted function and the distance of the start of the gene to the respective BSA peaks. Furthermore, T-DNA lines used for interspecific crosses are listed.

GeneID	Predicted Function (TAIR)	Distance FeiGa-Peak to Start (bp)	Distance SqGa-Peak to Start (bp)	Distance minor FeiLz-Peak to Start (bp)	T-DNA
AT5G61980	A member of ARF GAP domain (AGD)	100.555	88.818	49.914	SALK_036034
AT5G61990	Pentatricopeptide repeat (PPR) superfamily protein	95.421	83.684	44.780	NA
AT5G61997	This gene encodes a small protein and has either evidence of transcription or purifying selection	86.762	75.025	36.121	NA
AT5G62000	Encodes an auxin response factor	84.669	72.932	34.028	SALK_108995
AT5G62020	member of Heat Stress Transcription Factor (Hsf) family	78.938	67.201	28.297	SALK_012418
AT5G62030	diphthamide synthesis DPH2 family protein	76.548	64.811	25.907	GK-453F01-018814
AT5G62040	PEBP family protein	72.217	60.480	21.576	GK-331D12-016046
AT5G62050	essential factor for protein sorting and assembly into membranes	71.253	59.516	20.612	SALK_020304
AT5G62060	F-box and associated interaction domains-containing protein	67.339	55.602	16.698	SALK_077484
AT5G62065	Encodes a Protease inhibitor/seed storage/LTP family protein	65.820	54.083	15.179	GK-344E07-016232
AT5G62070	IQ-domain 23 (IQD23)	65.108	53.371	14.467	SAIL_823_B04
AT5G62080	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	61.556	49.819	10.915	NA
AT5G62090	SEUSS-like 2 (SLK2)	60.064	48.327	9.423	SALK_089954
AT5G62100	A member of Arabidopsis BAG (Bcl-2-associated athanogene) proteins	54.631	42.894	3.990	GK-493F03-019652
minor Fei-0 x Lz-0 Peak	**Position 24944386**	**50.641**	**38.904**	**0**	
AT5G62110	Homeodomain-like superfamily protein	52.858	41.121	2.217	NA
AT5G62120	member of Response Regulator: B- Type	48.072	36.335	-2.569	SALK_006711
AT5G62130	Per1-like family protein	44.763	33.026	-5.878	SALK_039375
AT5G62140	unknown protein	40.521	28.784	-10.120	SALK_008702
AT5G62150	peptidoglycan-binding LysM domain-containing protein	36.786	25.049	-13.855	SALK_144729
AT5G62160	member of Fe(II) transporter isolog family	34.920	23.183	-15.721	SALK_137184
AT5G62162	Encodes a phosphate starvation-responsive microRNA that targets PHO2	32.542	20.805	-18.099	NA
AT5G62165	Encodes a MADS box transcription factor	30.327	18.590	-20.314	GK-088H07-016184

AT5G62170	unknown protein	22.050	10.313	-28.591	SALK_047222
AT5G62180	carboxyesterase 20 (CXE20)	16.183	4.446	-34.458	SALK_043536
Sq-8 x Ga-0 Peak	**Position 24983290**	**11.737**	**0**	**-38.904**	
AT5G62190	DEAD/DEAH box RNA helicase PRH75	14.759	3.022	-35.882	SALK_040581
AT5G62200	Embryo-specific protein 3	10.791	-946	-39.850	SALK_129445
AT5G62210	Embryo-specific protein 3	8.700	-3.037	-41.941	SALK_038225
AT5G62220	Encodes a Golgi apparatus-localized galactosyltransferase.	6.571	-5.166	-44.070	GK-552C10-021687
Fei-0 x Ga-0 Peak	**Position 24995027**	**0**	**-11.737**	**-50.641**	
GWAS MLMM Peak	**Position 24996125**	**-1.098**	**-12.835**	**-51.739**	
AT5G62230	Encodes a receptor-like kinase	-1.102	-12.839	-51.743	SALK_084012C
AT5G62240	Cell cycle regulated microtubule associated protein	-7.331	-19.068	-57.972	SALK_099761
AT5G62250	microtubule-associated protein 65-9 (MAP65-9)	-10.729	-22.466	-61.370	GK-605A03-021275
AT5G62260	AT hook motif DNA-binding family protein	-14.000	-25.737	-64.641	SALK_061044
AT5G62270	unknown protein	-16.952	-28.689	-67.593	SALK_026223
AT5G62280	Protein of unknown function (DUF1442)	-22.046	-33.783	-72.687	SALK_095586
AT5G62290	nucleotide-sensitive chloride conductance regulator (ICln) family protein	-24.460	-36.197	-75.101	SALK_050231C
AT5G62300	Ribosomal protein S10p/S20e family protein	-26.173	-37.910	-76.814	WiscDsLox393-396H2
AT5G62310	Encodes a protein with a serine/threonine kinase domain	-28.328	-40.065	-78.969	NA
AT5G62320	Encodes a putative transcription factor (MYB99)	-33.689	-45.426	-84.330	SALK_003193
AT5G62330	Plant invertase/pectin methylesterase inhibitor superfamily protein	-36.574	-48.311	-87.215	SALK_092085
AT5G62340	Plant invertase/pectin methylesterase inhibitor superfamily protein	-38.491	-50.228	-89.132	NA
AT5G62350	Plant invertase/pectin methylesterase inhibitor superfamily protein	-42.399	-54.136	-93.040	NA
AT5G62360	Plant invertase/pectin methylesterase inhibitor superfamily protein	-45.622	-57.359	-96.263	SALK_073021
AT5G62370	Tetratricopeptide repeat (TPR)-like superfamily protein	-46.874	-58.611	-97.515	SAIL_890_F09

Table S6: Accession-specific reference genomes were constructed using publicly available SNP data and the Col-0 reference genome (www.arabidopsis.org) as a backbone.

Accession	Reference	SNP-file
Lz-0	*(12)*	cynin.gmi.oeaw.ac.at/home/resources/atpolydb/250k-snp-data/call_method_75.tar.gz
Fei-0	*(11)*	1001genomes.org/data/MPI/MPICao2010/releases/current/strains/Fei-0/TAIR10/filtered_variant.txt.gz
Ga-0	SALK, 1001genomes.org	1001genomes.org/data/Salk/releases/current/TAIR10/strains/Ga-0/quality_variant_filtered_Ga_0.txt.gz
Sq-8	SALK, 1001genomes.org	1001genomes.org/data/Salk/releases/current/TAIR10/strains/Sq-8/quality_variant_filtered_Sq_8.txt.gz

Table S7: Alignment Statistics. Reads of each pool were aligned to the reference genome of both parental accessions.

Initial Cross	Pool	Parent	Total Reads	Uniquely Aligned Reads	in Percent	Reads Failed To Align *	in Percent
Fei 0 x Lz 0	High	Fei 0	45916203	23177303	50.48%	4000865	8.71%
	Low	Fei 0	50199892	24877172	49.56%	4092869	8.15%
	High	Lz 0	45916203	20728035	45.14%	4006447	8.73%
	Low	Lz 0	50199892	22848127	45.51%	4099709	8.17%
Fei 0 x Ga 0	High	Fei 0	44362470	22696129	51.16%	3272208	7.38%
	Low	Fei 0	44760051	22182927	49.56%	3537092	7.90%
	High	Ga 0	44362470	22149706	49.93%	3282649	7.40%
	Low	Ga 0	44760051	22672548	50.65%	3548379	7.93%
Sq 8 x Ga 0	High	Sq 8	45535966	23401820	51.39%	3044846	6.69%
	Low	Sq 8	43894846	21781642	49.62%	3054449	6.96%
	High	Ga 0	45535966	22674974	49.80%	3044781	6.69%
	Low	Ga 0	43894846	22527179	51.32%	3057397	6.97%

* due to non uniqueness

Table S8: ***A. thaliana*** **accessions used for population genetic tests and the origin of SNP data.**

Accession	origin	Accession	origin
Ag-0	SALK	N13	SALK
An-1	SALK	Nd-1	CeBiT
Bay-0	JGI	NFA-10	SALK
Bil-5	GMI	NFA-8	SALK
Bil-7	GMI	Nok-3	SALK
Bor-1	SALK	Oemoe2-1	GMI
Bor-4	SALK	Oy-0	JGI
Br-0	SALK	Pna-10	SALK
Bur-0	MPI	Pna-17	SALK
C24	MPI	Pro-0	SALK
CIBC-17	SALK	Pu2-23	SALK
CIBC-5	SALK	Pu2-7	SALK
Col-0	MPI	Ra-0	SALK
Ct-1	WTC	Ren-1	SALK
Cvi-0	SALK	Ren-11	SALK
Eden-1	GMI	Rmx-A02	SALK
Eden-2	GMI	Rmx-A180	SALK
Edi-0	WTC	RRs-10	SALK
Ei-2	SALK	RRS-7	SALK
Est-1	MPI	Se-0	SALK
Faeb-2	GMI	Sha	MPI
Faeb-4	GMI	Sorbo	SALK
Fei-0	MPI	Sq-1	SALK
Ga-0	SALK	Sq-8	SALK
Got-22	SALK	Tamm-2	SALK
Got-7	SALK	Tamm-27	SALK
Gu-0	SALK	Ts-1	SALK
Gy-0	SALK	Ts-5	SALK
HR-10	SALK	Tsu-1	MPI
HR-5	SALK	Ull2-3	GMI
Kas-1	SALK	Ull2-5	GMI
Kin-0	SALK	Uod-1	SALK
Knox-18	SALK	Uod-7	SALK
Kondara	SALK	Vaar2-1	GMI
Kz-9	SALK	Van-0	SALK
Ler-1	SALK	Wa-1	SALK
love-1	GMI	Wei-0	SALK
love-5	GMI	Ws-0	WTC
Lp2-2	SALK	Ws-2	SALK
Lp2-6	SALK	Wt-5	SALK
Mt-0	WTC	Yo-0	SALK
Mz-0	SALK	Zdr-1	SALK

Chapter 5 – General Discussion

Arabidopsis as a model system to study speciation

Speciation genes drive the divergence of species by restricting gene flow between them (*1*). In our studies, we focused on the closely related species *A. thaliana* and *A. lyrata* as model systems in order to identify molecular factors establishing hybridization barriers in the Brassicaceae. The species diverged around 5 million years ago and are reproductively isolated mainly by their different mating systems: *A. thaliana* is predominantly selfing, whereas *A. lyrata* is an obligate self-incompatible out-crosser. Nevertheless, if *A. thaliana* is manually pollinated with *A. lyrata* pollen, additional incompatibilities are observed: *A. lyrata* pollen tubes (PTs) can grow normally through the *A. thaliana* transmitting tract and are properly guided towards the ovules. However, PT recognition and reception at the female gametophyte fails: instead of ceasing its growth and rupturing in order to release the sperm cells, the PT continues growing inside the female gametophyte, leaving it unfertilized (*2*). Although in nature chances are low that *A. lyrata* pollen reaches the stigma of *A. thaliana* due to their different mating systems, interspecific hybrids between the selfing *A. thaliana* and the out-crosser *A. arenosa* exist (*A. sueica* (*3*)). Furthermore, understanding the molecular basis of interspecific hybridization barriers in the genus Brassicaceae will give insights about the evolution of particular reproductive barriers and could be a pre-requisite for knowledge transfer into other species with potential applications in agriculture. In the future, it could possibly be used to develop tools in order to overcome existing crossing barriers. Furthermore, a multitude of genetic tools is available for *A. thaliana* including SNP data for thousands of accessions (*4*), enabling us to elucidate complex genetic traits underlying natural variation. In addition, the fact that a similar interspecific PT overgrowth phenotype as observed in crosses between *A. thaliana* and *A. lyrata* has been described in crosses of different Ericaceae species indicates that a conserved mechanism across plant families might control interspecific PT recognition (*2*, *5*, *6*).

Identification of *ARU* using GWAS

By crossing 86 *A. thaliana* accessions with interspecific *A. lyrata* pollen, we have shown that there is a large degree of natural variation in interspecific PT reception, depending on the mother plant. We found that some accessions, e.g. Lz-0, have the ability to almost

normally recognize and receive interspecific PTs from three Arabidopsis species (*A. lyrata*, *A. halleri* and *A. arenosa*) with only low proportions of ovules rejecting those PTs (10.20% of ovules with PT overgrowth per silique, OG/S). In contrast, almost all ovules of accessions like Kz-9 do not recognize and therefore reject interspecific PTs, leading to high proportions of ovules with PT overgrowth in a silique (87.30% OG/S for Kz-9). In order to identify genetic loci that are shaping this natural variation, we conducted a genome-wide association study (GWAS) with the variation in *A. lyrata* PT reception and publicly available SNP data for the 86 accessions (*7*).

GWAS is a powerful tool to correlate phenotypic variation with the underlying genetic basis and is widely used to identify susceptibility loci for complex human diseases and traits such as diabetes, cancer and height (reviewed in (*8*)). However, to date most of the identified SNPs were of minor effect and could only explain a small proportion of the phenotype. Most plant GWA studies have been conducted in *A. thaliana*, which favors association mapping by a high degree of homozygosity caused by selfing, as well as by the large amount of SNP-genotyped or sequenced accessions (*4*); but also crops like maize, rice and tomato have been used (*9*-*11*). In *Arabidopsis*, GWAS has been used in a proof-of-principle manner for several years to identify previously known candidate genes or to dissect the genetic architecture of a given trait (e.g. to understand the number of underlying loci and their relative contribution) (*7*, *12*-*17*). Only recently, novel genes have been identified with GWAS and were further confirmed using independent experiments (*18*, *19*). These genes are involved in *A. thaliana* root development and proline accumulation, respectively. With our GWAS, we identified a novel gene involved in the establishment of reproductive isolation barriers by influencing recognition and distinction of inter- and intraspecific PTs: *ARTUMES (ARU)*, coding for a subunit of the plant oligosaccharyltransferase (OST) complex, is regulating interspecific PT reception in the synergid cells of *A. thaliana*. However, in our case GWAS alone would not have resulted in the identification of a single causative gene. In order to narrow down the various GWAS peaks derived from General Linear Models (GLM), Mixed Linear Models (MLM) and Multi Locus Mixed Models (MLMM, see below), we tested mutants of candidate genes in these regions in interspecific crosses with *A. lyrata* pollen. If a gene is involved in interspecific PT reception, its mutant should behave differently from wild type. However, because most available T-DNA insertion lines are in the Col-0 background, with this strategy we would not identify genes that require a

genetic background different from Col-0 in order to show a phenotype. Despite that, among all 23 mutants tested, we could identify *aru*, which showed significantly higher OG/S in interspecific crosses than Col-0 wild type. Strengthening its role in the establishment of a hybridization barrier, the *ARU* locus shows signatures of positive selection (Fay and Wu's H) within *A. thaliana*, which is characteristic for rapidly evolving proteins shaping species evolution in animals (*20*).

We were able to identify *ARU* by conducting a GLM instead of the widely used MLM. Although we have carried out several MLM methods, including EMMA, EMMAX and P3D, they all resulted in singletons or peaks that were not significant (*21-23*). Despite that the GLM peak was also not significant using a 5% Bonferroni-corrected threshold, the peak on chromosome 1 at the position of the *ARU* locus was clearly higher than the others and included eight out of the top 20 most associated SNPs. Interestingly, MLM analysis on our phenotype data masked the peak. *A. thaliana* displays a strong population structure due to the lack of random mating, meaning that some accessions are more related to each other than to others (*24*). In other words, the genetic and phenotypic variances increase with geographic distance. Thus, linkage disequilibrium (LD, = the non-random co-occurrence between multiple alleles) is higher due to population structure and can cause synthetic associations and false positives (*25*). MLM takes population structure into account and reduces confounding effects of the genetic background; but it is also prone to create false negatives (SNPs are not detected although they account for the phenotype) (*26*), which is probably what happened in the *ARU* region when we used MLM. Most of the recent *A. thaliana* GWA studies use several hundreds of accessions to increase the statistical power (*18*, *19*). However, due to our time-consuming phenotyping approach consisting of emasculation, pollination with *A. lyrata* pollen, Aniline Blue staining and counting PT overgrowth at the fluorescence microscope, we chose a smaller number of 86 accessions and a potentially lower statistical power as a trade-off. Even with relatively few accessions, the variation of the observed phenotypes was maximal, ranging from about 10-90% OG/S. The maximal distribution of phenotypes might have favored the detection of a causative with only a small number of accessions.

In general, the genetic basis of traits that are determined by a small number of loci with relatively large effects is more favorable to be detected in GWAS than more complex traits. Such complex traits might be caused by multiple common variants with small

effects on a phenotype, or by the action of many rare variants of alleles with large effects. Thus, both the frequency of alleles in the sample and their relative effect on the phenotype are crucial for the power of GWAS to detect them (*25*). The gradual distribution of the phenotypic variation in our study indicates that the trait is determined by multiple loci with additive effects on the phenotype. Therefore it is likely that other genes in addition to *ARU* contribute to the variation in interspecific PT reception. In order to identify additional regulators and to overcome the obstacles posed by the small sample size and the tedious phenotyping, and to increase GWAS statistical power, two approaches are conceivable: (i) GWAS analysis with a stepwise MLMM, which includes the most significant SNP as a co-factor in the subsequent analysis steps. This allows the testing of multiple, dependent SNPs rather than single-locus tests like in classic approaches and was shown to outperform GLM and MLM both in simulations and in data sets from human and *Arabidopsis* (*27*). MLMM has been conducted on our data set, however we only analyzed a single T-DNA line in one of the genes that were near linked SNPs in step 5 (*At5g62230*), which did not show a phenotype in interspecific crosses (see Chapter 4). (ii) Dense sampling of local populations instead of the use of worldwide samples decreases heterogeneity (different alleles causing the same phenotype in different individuals) and population structure in the sample and can therefore be useful to identify otherwise rare alleles (*25*).

Unfortunately, we were not able to correlate the phenotypic variation observed in the accessions to sequence differences of ARU at the protein level, which could have explained altered functions. Since most of the variation in protein function in *A. thaliana* accessions is due to expression level differences (*28*), we tried to correlate *ARU* mRNA levels with extreme OG/S phenotypes. However, quantitative real-time (qRT) PCR analysis on RNA derived from emasculated pistils did not show any correlation of mRNA levels to the phenotypes. But since transgenic expression of *ARU* in *aru* synergid cells is sufficient to complement the mutant phenotype in interspecific crosses, *ARU* expression level differences only in these cells could be enough to generate the phenotypic variation we observed. These differences could be diluted when using mRNA from whole pistils and might therefore not be detectable in qRT PCR analysis. To solve this problem, qRT PCR analysis of synergid-derived RNA could be conducted, which depends on previous laser-capturing of synergids cells (*29*), a very time-consuming technique. We try to overcome this problem by measuring *ARU* mRNA levels in emasculated pistils using

digital droplet PCR (ddPCR; (*30*)). This method is much more sensitive than qRT PCR and might resolve minor differences in mRNA levels between the accessions that could explain the phenotypic variation. However, another critical point of linking *ARU* expression to phenotypic variation is the correct selection of accessions to include in the analysis. Due to the possible multigenic nature of the trait, extreme accessions might differ in other factors than *ARU* expression, an obstacle that could only be overcome by assessing expression level differences in all 86 accessions.

Glycosylation patterns determine species-specificity

The involvement of the OST subunit ARU in species-specific PT reception suggests that glycosylation patterns are a crucial factor for gametophyte interaction. OST is a key enzyme of N-glycosylation, as it is responsible for the transfer of a pre-assembled carbohydrate oligomer onto glycosylation sites of the nascent substrate protein at the ER membrane (*31*). ARU is homologous to the yeast subunits Ost3p and Ost6p, which were shown to be involved in the selection of specific glycosylation sites via their thioredoxin-like fold containing active-site cysteine residues in the luminal domain (*32*). However, ARU is lacking these active cysteines (*33*). Nevertheless, ARU seems to confer a certain degree of substrate specificity, as in *aru* mutants, only a subset of tested plasma membrane receptors remains un-glycosylated (*33*). The fact that specific OST subunits are required for the glycosylation of only a subset of proteins has also been observed in mammals. For example, knock-down of Ripophorin I (homologous to yeast Ost1p) seems to influence only the glycosylation of specific substrate proteins, but does not influence the general N-glycosylation machinery (*34*). In yeast, Ost3p and Ost6p define two alternative OST isoforms with distinct glycosylation specificity (*35*). One could imagine a scenario in plants, where ARU and its homolog OST3/6-LIKE (encoded by *At1g11560*) would act similarly. Thus it is possible that in synergid cells, ARU is mediating the specific glycosylation of a putative receptor protein that is distinguishing if an arriving PT originates from the same or from a different species (Figure 5.1). The putative receptor protein could act by binding to specific PT-derived signals, which would lead to a response reaction in the synergids, subsequently triggering PT burst and sperm release. Because *aru* mutants are not impaired in intraspecific reception of *A. thaliana* PTs, putative signals from these PTs must be perceived in a different way. Here,

receptor-ligand binding could mainly act via protein-protein interactions, which could be enhanced or facilitated by glycosylation of specific motives in the receptor but is not essential in the case of *A. thaliana* PTs. This would explain why we do not observe intraspecific PT overgrowth in *aru*. On the other hand, interspecific PT signals might be only poorly recognized by protein-protein binding, and the observed recognition of about 40% of *A. lyrata* PTs by *A. thaliana* wild-type synergids could be due to the presence of the carbohydrate decoration on the receptor conferring partial recognition. In *aru* synergids, however, mis- or hypo-glycosylation of the putative receptor could cause complete loss of interaction between the synergid receptor and its PT ligand leading to an almost complete rejection of interspecific PTs. A possible candidate for the synergid factor is the receptor-like kinase *FERONIA* due to the similarity of its mutant phenotype with interspecific PT overgrowth, however, direct evidence is lacking (see below).

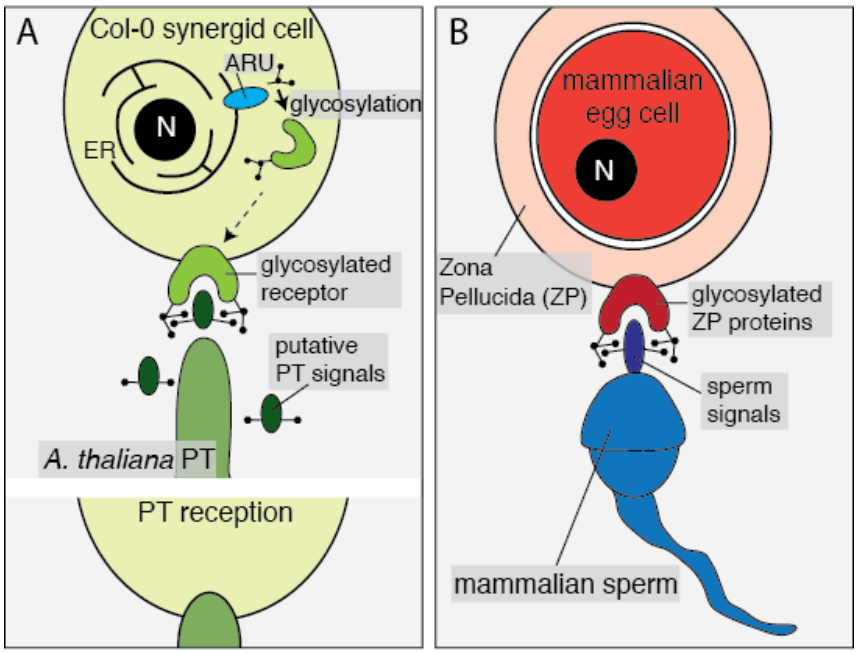


Figure 5.1: Pollen tube reception and mammalian fertilization share common characteristics. (**A**) Pollen tube reception in *A. thaliana* seems to depend on a putatively glycosylated synergid receptor interacting with unknown signals from the PT. The OST subunit ARU acting at the ER seems to play a role in species-specific glycosylation of the receptor. In an intraspecific cross (*A. thaliana* x *A. thaliana*), pollen signals might be recognized by the synergid receptor both via protein-protein and carbohydrate interactions. (**B**) Mammalian egg cells are surrounded by an extracellular matrix consisting of few types glycoproteins (zona pellucida, ZP). These highly glycosylated ZP proteins mediate initial sperm contact via carbohydrate and protein-protein binding; however the male components of this interaction remain elusive. ER: Endoplasmatic Reticulum, N: Nucleus.

Interestingly, the above described receptor-ligand binding by both protein-protein interactions and by interactions of carbohydrate residues has been observed during mammal fertilization (Figure 5.1). Here, the mouse oocyte is surrounded by a viscous extracellular matrix, the zona pellucida (ZP), which consists of three types of highly glycosylated proteins (ZP1-3). ZP3 is decorated with a variety of N- and O-linked glycosyl-residues and is critical for both initial sperm binding and triggering the acrosome reaction (*36*, *37*). For a long time, it was debated whether the initial ZP3-sperm binding is mediated by protein-protein or by carbohydrate interactions due to seemingly contradictory results favoring either one or the other mechanism. This lead to the development of the so-called domain-specific model, which proposed a critical role for both mechanisms (*38*). Only recently, there is raising evidence in human and mouse, that about 75% of sperm binding to the ZP3 is carbohydrate-dependent, whereas the remaining proportion binds via protein-protein interactions (*37*).

It has been shown that sperm-egg binding is a species-preferential process in many organisms (*39-41*). In sea invertebrates with external fertilization, where eggs and sperm are freely released into the water, species-specific recognition is crucial. Molecules on both the sperm and the egg have been identified that mediate gamete interaction: in sea urchins, the sperm adhesive protein bindin interacts with carbohydrates on a highly glycosylated egg surface receptor (egg bindin receptor 1, EBR1), which is thought to confer species-specific bindin-binding via the position of attachment of the carbohydrate residues determined by the protein sequence (*41-44*); and in abalone, the sperm protein lysin binds to the egg vitelline envelope receptor for lysin (VERL) in a lectin-like manner (*40*). Both receptor-ligand interactions are species-specific, allowing successful fertilization even if sperm and eggs of different species are mixed in the sea due to overlapping habitats and breeding seasons. The situation is a bit more relaxed in mammals (and plants), which have internal fertilization, where several layers of prior species barriers have to be overcome before egg and sperm actually interact. However, also in mammals, sperm-ZP binding is species-specific (*39*). For example, human spermatoids are able to bind to human and gibbon oocytes *in vitro*, but not to oocytes from non-hominoid primates or even more distant species like mouse or pig. Interestingly, sperm derived from other mammals like rabbit or mouse did not exhibit such specificity *in vitro* and even bound to human oocytes (*39*).

Similarly to the situation during animal fertilization, our findings suggest that (species-specific) glycosylation patters regulate PT-embryo sac interaction in plants, although so far, neither the nature of the PT signals nor the synergid receptors are known. However, N-glycosylation has not only been proven to be important in interspecific PT reception by the isolation of *ARU*, but also in intraspecific PT reception N-glycosylation seems to play a role: mutants in the UDP-glycosyltransferase *TURAN* (*TUN*) show reduced PT reception and PT overgrowth (*45*). ARU and TUN are constituents of the same N-glycosylation pathway at the ER membrane. TUN acts in early stages of the N-glycosylation process during the assembly of the carbohydrate oligomer and therefore might have a broader effect on the glycoproteome compared to ARU, which only seems to influence specific glycoproteins (*33*). This could explain why *TUN* has impacts on the general PT reception pathway and shows a phenotype also in intraspecific pollinations, whereas *ARU* only influences interspecific PT reception.

FERONIA is likely not involved in interspecific PT recognition

Because interspecific crosses between Rhododendron species and various species within the Brassicaceae family resulted in *feronia*-like PT overgrowth in the ovules, and because the extracellular domain of *FERONIA* (*FER*) displayed signatures of rapid evolution in pairwise comparisons between Brassicaceae species, it was proposed that FER directly mediates the recognition and rejection of interspecific PTs (*2*, *6*). Among the tested interspecific crosses was *A. thaliana* pollinated by *A. lyrata*, resulting in more than 50% of ovules with PT overgrowth in a silique (*2*). However, expression of the *FER*-version derived from *A. lyrata* in *A. thaliana* homozygous *fer* mutants did not improve the recognition and reception of *A. lyrata* PTs. In addition, these transgenic ovules were perfectly able to receive *A. thaliana* PTs. These results indicate that the peptide sequence of *FER* is not directly involved in the species-specific binding of putative ligands of interspecific PTs. Only recently, the ligand of FER in roots has been identified to be a rapid alkalization factor (RALF), which was shown to bind to the protein sequence of FER, resulting in its phosphorylation and activation (*46*). In *A. thaliana*, RALF has 34 homologs (RALF-LIKE peptides), several of which are expressed in pollen, thus representing good candidates for interacting with FER in the synergids during PT reception. However, FER possesses eight putative glycosylation sites in the extracellular domain and carries two putative carbohydrate-binding malectin-like domains (*47*, *48*),

suggesting that FER might interact with glycans. At least some of its glycosylation sites seem to be occupied by carbohydrate residues, as it was shown that enzymatic de-glycosylation of FER resulted in a band-shift in western-blot analysis compared to the untreated FER protein. In addition, the identification of *ARU* hints that glycosylation patterns confer species-specificity in PT reception, although FER does not seem to be a direct target of ARU as *aru* mutants are not impaired in intraspecific PT reception and FER-GFP is normally localized in *aru* synergids.

However, in the light of the importance of glycosylation for interspecific PT reception, one could imagine that in *fer* mutants (in the *A. thaliana* L*er* background), the above-mentioned, transgenic-expressed *A. lyrata* FER might be decorated with *A. thaliana*-specific glycosylation patterns and could therefore function similar to FER of *A. thaliana*. This specific *A. thaliana* glycosylation pattern of the *A. lyrata* FER protein might at the same time prevent better recognition of the putative *A. lyrata* PT ligand in interspecific crosses. A similar phenomenon has been observed in mouse oocytes expressing human ZP3, which acquires mouse-specific O-glycans and does not bind human, but murine sperm (*49*). This hypothesis, however, is speculative and based on the assumption that FER in synergids binds its ligand solely via carbohydrate interactions, which is rather unlikely in the light of the RALF-FERONIA ligand-receptor binding conferred by protein-protein interactions in roots (*46*). However, it is possible that FER might have additional ligands interacting with its carbohydrate residues. Further studies on the glycosylation status of FER and pollen-expressed RALF-LIKE peptides as possible ligands will shed more light on this process.

Identification of *At5g62150* using Bulk-Segregant Analysis

We have shown that *A. thaliana* accessions differ in their ability to recognize interspecific PTs, resulting in differing proportions of OG/S. We have generated hybrid F_2 populations between accessions with very low and very high OG/S, enabling us to determine the number of loci defining the trait between a pair of accessions and opens up the possibility to map regulators with a very strong effect on interspecific PT recognition. In contrast to GWAS, which takes the diversity of large numbers of accessions into account, this approach is limited to the allelic diversity between two parental accessions and strongly depends on the recombination rate within one

generation (from F_1 to F_2). However, such an F_2 mapping approach reduces the complexity of the trait compared to GWAS, making it more feasible to identify single genes with large effects.

We analyzed the segregation of individuals with low, intermediate and high OG/S in a total of eight F_2 populations derived from different combinations of parental accessions. Five of them showed a gradual distribution of phenotypes indicating that multiple loci linked to the phenotype are segregating within the population. The three remaining populations showed a segregation ratio of 1:2:1 (low:intermediate:high OG/S), suggesting that between the respective parental accession pairs, a single locus is determining the trait. The respective accession pairs were Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0. Although it would theoretically be also possible to map the genetic basis of the trait variation in the F_2 populations with a gradual distribution (*50*), we focused on the three populations with 1:2:1 segregation ratio in order to map a single gene with maximal influence on the trait. Allelic variation in this gene would be causing either around 10% (low accession) or 80% (high accession) of ovules within a silique to reject interspecific PTs. For each population, we pooled tissue derived from F_2 individuals with either very low or very high OG/S separately and subjected the DNA to Next-Generation Illumina HiSeq sequencing. For each low and high pool, the reads were mapped to genomes of both parental accessions. We expected that in each pool, causative alleles linked to the trait would be derived from one parent only (allele frequency of 1), whereas unlinked SNPs would segregate randomly in each pool (allele frequency of 0.5) (*50*). Using a modified G' statistic for Bulk-Segregant Analysis (BSA) that takes allele frequencies into account (*51*, *52*), we were able to identify the same region on chromosome 5 for all three accession pairs (Fei-0 x Ga-0, Sq-8 x Ga-0 and Fei-0 x Lz-0), suggesting a major regulator for interspecific PT recognition in that region. Because initially at least one of the parental accessions Fei-0 or Ga-0 was used in all crosses, it is not surprising that we detect the same region with all three F_2 populations. The BSA peaks of both F_2 populations with Ga-0 as a low parent (Fei-0 x Ga-0, Sq-8 x Ga-0) were in very close proximity (only around 12 kbp apart). Interestingly, the exact same region was also identified with GWAS MLMM (step 5) on OG/S of 86 accessions obtained in our initial phenotyping experiment. In contrast, the main peak of Fei-0 x Lz-0 was about 1.1 Mbp away. In addition, for this accession combination several smaller peaks were detected, one of which located around the region of the Fei-0 x Ga-0 and Sq-8 x Ga-0

maxima. There are several possible explanations for this result: i) Between Fei-0 and Lz-0, two (or more) very closely linked loci are responsible for the variation in the trait; ii) an inversion in the Lz-0 genome could cause this mapping bias; iii) the resolution might be impaired due to the different source of the Lz-0 SNP data compared to the other parental accessions (Lz-0 data from (*53*), others from 1001 genomes: (*4*). Focusing on a 150 kb candidate region determined by the BSA peaks, we conducted a reverse genetics screen with T-DNA insertion lines. A T-DNA insertion line disrupting the gene *At5g62150* showed significantly higher proportions of OG/S than Col-0 wild type, making *At5g62150* a likely candidate involved in interspecific PT reception. This will be further confirmed with an additional mutant allele and by complementation of the mutant phenotype with functional *At5g62150* alleles of low and high accessions. The fact that the coding sequence of *At5g62150* does not contain stop-codons or other amino acid exchanges in the alleles of the parental accessions suggests that expression level differences rather than a complete protein loss-of-function might cause the phenotypic variation in parental accessions. We will address this question by ddPCR analysis of *At5g62150* expression levels in emasculated pistils (*30*).

The LysM domain and its potential role in PT reception

At5g62150 codes for a small peptide containing a glycan-binding LysM domain. The Lysin motif (LysM) has first been described in the lysozyme of *Bacillus* phage ϕ29 and the *Enterococcus faecalis* peptidoglycan hydrolase (*54*, *55*). In bacteria and phage, the motif is usually coupled to glycan-degrading enzymes. Thus, LysM glycan-binding activity directs the enzymes to their targets. In contrast, in plants, all LysM-domain containing proteins so far identified are extracellular receptors with or without an intracellular kinase domain, which directly or indirectly mediate immune response after pathogen infection (both bacterial and fungal invasion) (*56-59*), or are implicated in the establishment of symbiosis (arbuscular mycorrhiza formation with Glomeromyceta as well as root nodule formation with Rhizobacteria) (*60*, *61*). Interestingly, PT reception and plant defense have been previously shown to share common molecular components: *NTA*, a member of the MLO protein family involved in powdery mildew susceptibility, regulates PT reception (*62*). In addition, *FER*, the first identified gene to be involved in PT reception, seems to play a role during fungal invasion, as *fer* mutants are resistant to powdery mildew.

Plant LysM domain proteins are grouped into four categories: plasma membrane localized LysM receptor kinases, membrane-anchored proteins with an extracellular LysM domain but without an intracellular kinase domain, as well as two groups of small peptides with only one Lys motif that can either be secreted or remain intracellularly (*63*). *At5g62150* codes for a small peptide with only one LysM, likely belonging to one of the two latter groups. However, so far it remains speculative if At5g62150 is secreted or not and microscopic analysis of the fluorescence-tagged protein will help to answer this question. The so far described plant LysM proteins have extracellular Lysin motives directly binding GlcNAc (N-acetylglucosamine) residues derived from bacterial cell wall peptidoglycans or chitin, a major constituent of fungal cell walls (*59*, *64*). In plants, GlcNAc is a common posttranslational modification of glycoproteins and is attached during N- and O-glycosylation (*65*, *66*). During N-glycosylation, GlcNAc is part of the pre-assembled core carbohydrate oligomer that is transferred to the nascent polypeptide by OST. Later, during glycoprotein modification in the ER and Golgi, even more GlcNAc residues can be attached (*65*). In our GWAS, we have identified the OST-subunit *ARU* to be a regulator of interspecific PT reception. Consequently, given the identification of both an enzyme involved in protein glycosylation and a small peptide that possibly directly binds such glycoproteins, the involvement of glycosylation patterns in species-specific PT recognition is even more strengthened. It is conceivable that At5g62150 binds to a putative receptor glycoprotein on the PT in a species-specific manner, triggering PT growth arrest and burst. Furthermore, it is possible that At5g62150 forms an active complex with a putative synergid receptor, leading to species-specific binding of carbohydrate signals from the arriving PT. Further characterization of At5g62150 protein function will help to understand the molecular processes underlying the discrimination of intra- and interspecific PTs by the female gametophyte.

References

1. L. H. Rieseberg, B. K. Blackman, Speciation genes in plants, *Annals of botany* **106**, 439–455 (2010).

2. J. M. Escobar-Restrepo *et al.*, The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception, *Science* **317**, 656–660 (2007).

3. M. Jakobsson *et al.*, A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers, *Molecular Biology and Evolution* **23**, 1217–1231 (2006).

4. J. Cao *et al.*, Whole-genome sequencing of multiple *Arabidopsis thaliana* populations, *Nature genetics* **43**, 956–963 (2011).

5. V. R. J. L. W. E. G. Kaul, Early events in the embryo sac after intraspecific and interspecific pollinations in *Rhododendron kawakamii* and *R. retusum*, *Canadian Journal of Botany* **64**, 282–291 (1986).

6. E. G. Williams, V. Kaul, J. L. Rouse, B. F. Palser, Overgrowth of Pollen Tubes in Embryo Sacs of *Rhododendron* Following Interspecific Pollinations, *Aust. J. Bot.* **34**, 413–423 (1986).

7. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, *Nature* **465**, 627–631 (2010).

8. L. Hou, H. Zhao, A review of post-GWAS prioritization approaches, *Front Genet* **4**, 280 (2013).

9. F. Tian *et al.*, Genome-wide association study of leaf architecture in the maize nested association mapping population, *Nature genetics* **43**, 159–162 (2011).

10. X. Huang *et al.*, Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm, *Nature genetics* **44**, 32–39 (2012).

11. N. Ranc *et al.*, Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*, *G3 (Bethesda)* **2**, 853–864 (2012).

12. A. Nemri *et al.*, Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping, *Proceedings of the National Academy of Sciences* **107**, 10302–10307 (2010).

13. I. Baxter *et al.*, A Coastal Cline in Sodium Accumulation in *Arabidopsis thaliana* Is Driven by Natural Variation of the Sodium Transporter *AtHKT1;1, PLoS Genet* **6**, e1001193 (2010).

14. Y. Li, Y. Huang, J. Bergelson, M. Nordborg, J. O. Borevitz, Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*, *Proceedings of the National Academy of Sciences* **107**, 21199–21204 (2010).

15. D. L. Filiault, J. N. Maloof, A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response, *PLoS Genet* **8**, e1002589 (2012).

16. D.-Y. Chao *et al.*, Genome-wide association studies identify heavy metal *ATPase3* as the primary determinant of natural variation in leaf cadmium in *Arabidopsis thaliana*, *PLoS Genet* **8**, e1002923 (2012).

17. Y. Li, R. Cheng, K. A. Spokas, A. A. Palmer, J. O. Borevitz, Genetic Variation for Life History Sensitivity to Seasonal Warming in *Arabidopsis thaliana*, *Genetics* **196**, 569–577 (2014).

18. M. Meijon, S. B. Satbhai, T. Tsuchimatsu, W. Busch, Genome-wide association study using cellular traits identifies a new regulator of root development in *Arabidopsis*, *Nature genetics* (2013).

19. P. E. Verslues, J. R. Lasky, T. E. Juenger, T. W. Liu, M. N. Kumar, Genome wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis thaliana*, *Plant physiology* (2013).

20. V. D. Vacquier, W. J. Swanson, Selection in the rapid evolution of gamete recognition proteins in marine invertebrates, *Cold Spring Harb Perspect Biol* **3** (2011).

21. H. M. Kang *et al.*, Efficient control of population structure in model organism association mapping, *Genetics* **178**, 1709–1723 (2008).

22. H. M. Kang *et al.*, Variance component model to account for sample structure in genome-wide association studies, *Nature genetics* **42**, 348–354 (2010).

23. Z. Zhang *et al.*, Mixed linear model approach adapted for genome-wide association studies, *Nature genetics* **42**, 355–360 (2010).

24. A. Platt *et al.*, The scale of population structure in *Arabidopsis thaliana*, *PLoS Genet* **6**, e1000843 (2010).

25. A. Korte, A. Farlow, The advantages and limitations of trait analysis with GWAS: a review, *Plant Methods* **9**, 29 (2013).

26. J. Bergelson, F. Roux, Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*, *Nature reviews. Genetics* **11**, 867–879 (2010).

27. V. Segura *et al.*, An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, *Nature genetics* **44**, 825–830 (2012).

28. X. Gan *et al.*, Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*, *Nature* **477**, 419–423 (2011).

29. R. C. Day, U. Grossniklaus, R. C. Macknight, Be more specific! Laser-assisted microdissection of plant cells, *Trends Plant Sci.* **10**, 397–406 (2005).

30. B. J. Hindson *et al.*, High-throughput droplet digital PCR system for absolute quantitation of DNA copy number, *Anal. Chem.* **83**, 8604–8610 (2011).

31. M. Aebi, N-linked protein glycosylation in the ER, *Biochimica et biophysica acta* **1833**, 2430–2437 (2013).

32. B. L. Schulz *et al.*, Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency, *Proceedings of the National Academy of Sciences* **106**, 11061–11066 (2009).

33. A. Farid *et al.*, Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses, *Plant physiology* **162**, 24–38 (2013).

34. C. M. Wilson, S. High, Ribophorin I acts as a substrate-specific facilitator of N-glycosylation, *J. Cell. Sci.* **120**, 648–657 (2007).

35. M. Schwarz, R. Knauer, L. Lehle, Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit, *FEBS letters* **579**, 6564–6568 (2005).

36. J. D. Bleil, P. M. Wassarman, Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm, *Cell* **20**, 873–882 (1980).

37. G. F. Clark, A role for carbohydrate recognition in mammalian sperm-egg binding, *Biochem. Biophys. Res. Commun.* (2014).

38. G. F. Clark, Molecular models for mouse sperm-oocyte binding, *Glycobiology* **21**, 3–5 (2011).

39. J. M. Bedford, Sperm/egg interaction: the specificity of human spermatozoa, *Anat. Rec.* **188**, 477–487 (1977).

40. W. J. Swanson, V. D. Vacquier, The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule, *Proceedings of the National Academy of Sciences* **94**, 6724–6729 (1997).

41. N. Kamei, C. G. Glabe, The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein, *Genes & Development* **17**, 2502–2507 (2003).

42. C. G. Glabe, V. D. Vacquier, Egg surface glycoprotein receptor for sea urchin sperm bindin, *Proceedings of the National Academy of Sciences* **75**, 881–885 (1978).

43. C. G. Glabe, L. B. Grabel, V. D. Vacquier, S. D. Rosen, Carbohydrate specificity of sea urchin sperm bindin: a cell surface lectin mediating sperm-egg adhesion, *J. Cell Biol.* **94**, 123–128 (1982).

44. N. Hirohashi, W. J. Lennarz, Role of a vitelline layer-associated 350 kDa glycoprotein in controlling species-specific gamete interaction in the sea urchin, *Dev. Growth Differ.* **43**, 247–255 (2001).

45. H. Lindner *et al.*, SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing, *Genetics* (2012).

46. M. Haruta, G. Sabat, K. Stecker, B. B. Minkoff, M. R. Sussman, A peptide hormone and its receptor protein kinase regulate plant cell expansion, *Science* **343**, 408–411 (2014).

47. H. Lindner, L. M. Muller, A. Boisson-Dernier, U. Grossniklaus, *Cr*RLK1L receptor-like kinases: not just another brick in the wall, *Current opinion in plant biology* **15**, 659–669 (2012).

48. A. Boisson-Dernier, S. A. Kessler, U. Grossniklaus, The walls have ears: the role of plant *Cr*RLK1Ls in sensing and transducing extracellular signals, *Journal of Experimental Botany* **62**, 1581–1591 (2011).

49. A. Dell *et al.*, Murine and human zona pellucida 3 derived from mouse eggs express identical O-glycans, *Proceedings of the National Academy of Sciences* **100**, 15631–15636 (2003).

50. H. Takagi *et al.*, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, *The Plant Journal* **74**, 174–183 (2013).

51. P. M. Magwene, J. H. Willis, J. K. Kelly, The statistics of bulk segregant analysis using next generation sequencing, *PLoS Comput. Biol.* **7**.

52. Z. Yang *et al.*, Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes, *PLoS ONE* **8**, e68433 (2013).

53. M. W. Horton *et al.*, Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel, *Nature genetics* **44**, 212–216 (2012).

54. K. J. Garvey, M. S. Saedi, J. Ito, Nucleotide sequence of Bacillus phage phi 29 genes 14 and 15: homology of gene 15 with other phage lysozymes, *Nucleic Acids Res.* **14**, 10001–10008 (1986).

55. C. Béliveau, C. Potvin, J. Trudel, A. Asselin, G. Bellemare, Cloning, sequencing, and expression in *Escherichia coli* of a *Streptococcus faecalis* autolysin, *J. Bacteriol.* **173**, 5619–5623 (1991).

56. H. Kaku *et al.*, Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor, *Proceedings of the National Academy of Sciences* **103**, 11086–11091 (2006).

57. A. Miya *et al.*, CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*, *Proceedings of the National Academy of Sciences* **104**, 19613–19618 (2007).

58. T. Shimizu *et al.*, Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice, *The Plant Journal* **64**, 204–214 (2010).

59. R. Willmann *et al.*, *Arabidopsis* lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection, *Proceedings of the National Academy of Sciences* **108**, 19824–19829 (2011).

60. S. Radutoiu *et al.*, Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases, *Nature* **425**, 585–592 (2003).

61. E. Limpens *et al.*, LysM domain receptor kinases regulating rhizobial Nod factor-induced infection, *Science* **302**, 630–633 (2003).

62. S. A. Kessler *et al.*, Conserved molecular components for pollen tube reception and fungal invasion, *Science* **330**, 968–971 (2010).

63. A. A. Gust, R. Willmann, Y. Desaki, H. M. Grabherr, T. Nürnberger, Plant LysM proteins: modules mediating symbiosis and immunity, *Trends Plant Sci.* **17**, 495–502 (2012).

64. E. K. Petutschnig, A. M. E. Jones, L. Serazetdinova, U. Lipka, V. Lipka, The lysin motif receptor-like kinase (LysM-RLK) CERK1 is a major chitin-binding protein in *Arabidopsis thaliana* and subject to chitin-induced phosphorylation, *The Journal of biological chemistry* **285**, 28902–28911 (2010).

65. I. B. H. Wilson, Glycosylation of proteins in plants and invertebrates, *Curr. Opin. Struct. Biol.* **12**, 569–577 (2002).

66. N. E. Olszewski, C. M. West, S. O. Sassi, L. M. Hartweck, O-GlcNAc protein modification in plants: Evolution and function, *Biochimica et biophysica acta* **1800**, 49–56 (2010).

Appendix

Contents

- GWAS MLM results (OG/S): Tables A1 and A2, referred to in Chapter 3 (additional results)
- GWAS MLM results (Ovules not attracting pollen tubes): Tables A3 and A4, referred to in Chapter 3 (additional results)
- Vector Maps: referred to in Chapters 2 and 3

Table A1: Top 100 SNPs for OG/S, full SNP set (TASSEL P3D). Chr: Chromosome

rank	Marker	Chr	Site	p
1	PERL0617038	3	18010190	2,33E-06
2	PERL0901506	5	6068770	6,50E-05
3	PERL0667390	4	979996	6,54E-05
4	PERL0382632	2	14123516	7,03E-05
5	PERL0000734	1	119003	7,70E-05
6	PERL0000813	1	133186	8,14E-05
7	PERL0000822	1	133626	8,14E-05
8	PERL0000823	1	133701	8,14E-05
9	PERL0000832	1	134063	8,14E-05
10	PERL0891072	5	4899000	8,29E-05
11	PERL1043650	5	18084064	8,44E-05
12	PERL0219430	1	24325910	9,21E-05
13	PERL0891205	5	4924090	9,37E-05
14	PERL0183896	1	21605799	1,03E-04
15	PERL0759382	4	8499071	1,08E-04
16	PERL0759366	4	8497876	1,37E-04
17	PERL0759396	4	8501077	1,37E-04
18	PERL0759416	4	8502544	1,37E-04
19	PERL0759533	4	8507406	1,61E-04
20	PERL0278005	2	1043963	2,08E-04
21	PERL0392478	2	15995411	2,08E-04
22	PERL1095958	5	22934169	2,22E-04
23	PERL1022663	5	16827147	2,40E-04
24	PERL0759481	4	8504510	2,76E-04
25	PERL0062096	1	8320546	2,79E-04
26	PERL0384681	2	14516942	2,84E-04
27	PERL0401639	2	17494583	2,90E-04
28	PERL0863071	5	944253	2,91E-04
29	PERL0061956	1	8307842	3,01E-04
30	PERL0393659	2	16248770	3,04E-04
31	PERL0998055	5	15096264	3,16E-04
32	PERL0667417	4	982814	3,29E-04
33	PERL0668831	4	1113277	3,85E-04
34	PERL0161383	1	18845465	3,88E-04
35	PERL0621038	3	18393298	4,00E-04
36	PERL1070109	5	19994195	4,18E-04
37	PERL1014570	5	16250095	4,25E-04
38	PERL0112360	1	13117521	4,30E-04
39	PERL0667367	4	978836	4,33E-04
40	PERL0726098	4	6430376	4,33E-04
41	PERL0759524	4	8507015	4,37E-04
42	PERL0000724	1	118192	4,43E-04
43	PERL0759400	4	8501606	4,44E-04
44	PERL0726223	4	6441807	4,67E-04
45	PERL0440595	3	3394526	4,82E-04
46	PERL0278488	2	1103570	4,91E-04
47	PERL0615461	3	17867164	4,94E-04
48	PERL0059088	1	8004390	5,08E-04
49	PERL0891074	5	4899062	5,09E-04

50	PERL0760391	4	8549195	5,19E-04
51	PERL1028346	5	17240670	5,26E-04
52	PERL0166252	1	19503734	5,40E-04
53	PERL1044452	5	18140461	5,54E-04
54	PERL0667380	4	979449	5,64E-04
55	PERL0661920	4	505633	5,72E-04
56	PERL0387191	2	15019298	5,73E-04
57	PERL0440663	3	3399660	5,83E-04
58	PERL0318568	2	5315357	5,98E-04
59	PERL0277975	2	1040516	5,99E-04
60	PERL0697601	4	3798291	6,15E-04
61	PERL0009578	1	1476349	6,15E-04
62	PERL1095880	5	22923362	6,18E-04
63	PERL0710741	4	5187362	6,24E-04
64	PERL0667242	4	972411	6,29E-04
65	PERL0291095	2	2220611	6,52E-04
66	PERL0620601	3	18349378	6,55E-04
67	PERL0130427	1	15974746	6,65E-04
68	PERL0112362	1	13117647	6,79E-04
69	PERL0695041	4	3511253	6,83E-04
70	PERL0354981	2	9749374	6,90E-04
71	PERL0668950	4	1117210	7,09E-04
72	PERL0886001	5	4046208	7,57E-04
73	PERL0383164	2	14222645	7,65E-04
74	PERL0183402	1	21574903	7,69E-04
75	PERL0048266	1	6593084	7,83E-04
76	PERL0357890	2	10278238	7,89E-04
77	PERL0450010	3	4596537	7,92E-04
78	PERL0612470	3	17547736	7,97E-04
79	PERL0837159	4	16261566	7,98E-04
80	PERL0837161	4	16261800	7,98E-04
81	PERL0000616	1	105029	8,06E-04
82	PERL0060539	1	8163595	8,06E-04
83	PERL0264237	1	29804689	8,06E-04
84	PERL0273693	2	636392	8,06E-04
85	PERL0084889	1	11015568	8,11E-04
86	PERL0339587	2	7531633	8,31E-04
87	PERL0339588	2	7531688	8,31E-04
88	PERL0264399	1	29815614	8,41E-04
89	PERL0397350	2	16883935	8,49E-04
90	PERL0277871	2	1031490	8,62E-04
91	PERL0277934	2	1036991	8,62E-04
92	PERL0026942	1	3948738	9,12E-04
93	PERL0156060	1	18133466	9,14E-04
94	PERL0410739	2	18734611	9,16E-04
95	PERL0005899	1	844884	9,27E-04
96	PERL1014968	5	16276663	9,28E-04
97	PERL1028324	5	17239650	9,40E-04
98	PERL1085347	5	21629772	9,46E-04
99	PERL1085368	5	21637711	9,46E-04
100	PERL0934541	5	9143921	9,52E-04

Table A2: Top 100 SNPs for OG/S, SNP0.1 (TASSEL P3D). Chr: Chromosome

rank	Marker	Chr	Site	p
1	PERL0617038	3	18010190	1,82E-06
2	PERL0667390	4	979996	4,95E-05
3	PERL0000734	1	119003	5,81E-05
4	PERL0382632	2	14123516	6,45E-05
5	PERL0219430	1	24325910	7,86E-05
6	PERL1043650	5	18084064	8,12E-05
7	PERL0901506	5	6068770	1,48E-04
8	PERL0000813	1	133186	1,68E-04
9	PERL0000822	1	133626	1,68E-04
10	PERL0000823	1	133701	1,68E-04
11	PERL0000832	1	134063	1,68E-04
12	PERL0667236	4	972178	1,72E-04
13	PERL0183896	1	21605799	1,90E-04
14	PERL1022663	5	16827147	2,20E-04
15	PERL0759481	4	8504510	2,27E-04
16	PERL1095958	5	22934169	2,40E-04
17	PERL0401639	2	17494583	2,69E-04
18	PERL0621038	3	18393298	2,85E-04
19	PERL0998055	5	15096264	3,24E-04
20	PERL0667380	4	979449	3,41E-04
21	PERL0392478	2	15995411	3,43E-04
22	PERL0667367	4	978836	3,45E-04
23	PERL0000724	1	118192	3,50E-04
24	PERL0062096	1	8320546	3,53E-04
25	PERL0440663	3	3399660	3,70E-04
26	PERL0278005	2	1043963	3,83E-04
27	PERL0668831	4	1113277	3,91E-04
28	PERL0620601	3	18349378	3,95E-04
29	PERL1014570	5	16250095	3,96E-04
30	PERL0726223	4	6441807	4,27E-04
31	PERL1070109	5	19994195	4,32E-04
32	PERL0059088	1	8004390	4,54E-04
33	PERL0112360	1	13117521	4,60E-04
34	PERL0339587	2	7531633	4,60E-04
35	PERL0339588	2	7531688	4,60E-04
36	PERL0760391	4	8549195	4,71E-04
37	PERL0661920	4	505633	4,83E-04
38	PERL0863071	5	944253	5,26E-04
39	PERL0615461	3	17867164	5,28E-04
40	PERL1028346	5	17240670	5,39E-04
41	PERL0166252	1	19503734	5,63E-04
42	PERL0710741	4	5187362	5,66E-04
43	PERL0329435	2	6392760	5,77E-04
44	PERL0112362	1	13117647	5,78E-04
45	PERL0387191	2	15019298	5,81E-04
46	PERL0357890	2	10278238	5,87E-04
47	PERL0837159	4	16261566	5,92E-04
48	PERL0837161	4	16261800	5,92E-04
49	PERL0695041	4	3511253	6,13E-04

50	PERL0278488	2	1103570	6,29E-04
51	PERL0130427	1	15974746	6,35E-04
52	PERL0264399	1	29815614	6,56E-04
53	PERL1095880	5	22923362	6,67E-04
54	PERL0383164	2	14222645	6,78E-04
55	PERL0667417	4	982814	6,79E-04
56	PERL0009578	1	1476349	6,92E-04
57	PERL0084889	1	11015568	7,01E-04
58	PERL0291095	2	2220611	7,07E-04
59	PERL0726098	4	6430376	7,11E-04
60	PERL0410739	2	18734611	7,18E-04
61	PERL0668950	4	1117210	7,38E-04
62	PERL0657629	4	2876	7,40E-04
63	PERL0886001	5	4046208	7,58E-04
64	PERL0339577	2	7528705	7,73E-04
65	PERL0345691	2	8285685	7,75E-04
66	PERL0397350	2	16883935	7,83E-04
67	PERL1096811	5	23079647	7,96E-04
68	PERL1014988	5	16277911	7,98E-04
69	PERL1014989	5	16277937	7,98E-04
70	PERL0612470	3	17547736	8,45E-04
71	PERL0005899	1	844884	8,51E-04
72	PERL0354981	2	9749374	8,70E-04
73	PERL0279067	2	1177488	8,73E-04
74	PERL1014968	5	16276663	8,75E-04
75	PERL0277975	2	1040516	8,79E-04
76	PERL1085347	5	21629772	9,10E-04
77	PERL1085368	5	21637711	9,10E-04
78	PERL0796368	4	11002939	9,15E-04
79	PERL0836944	4	16247716	9,38E-04
80	PERL0615787	3	17891426	9,44E-04
81	PERL0836967	4	16249593	9,45E-04
82	PERL0836950	4	16248630	9,73E-04
83	PERL0918441	5	7795574	9,77E-04
84	PERL0996614	5	15041429	0,001016975
85	PERL0013784	1	2137228	0,001033176
86	PERL0754190	4	8129403	0,001033224
87	PERL1014086	5	16207512	0,001037331
88	PERL1014089	5	16207742	0,001037331
89	PERL0620445	3	18332805	0,001037657
90	PERL0969462	5	12557315	0,001054206
91	PERL0318568	2	5315357	0,001066695
92	PERL1096873	5	23086332	0,001077209
93	PERL1028324	5	17239650	0,001117363
94	PERL1095785	5	22901077	0,001130139
95	PERL1097478	5	23184170	0,001135142
96	PERL0156060	1	18133466	0,001145894
97	PERL0853591	5	58898	0,001162489
98	PERL1071398	5	20140364	0,001181062
99	PERL1071400	5	20140702	0,001181062
100	PERL1118073	5	25993535	0,001191786

Table A3: Top 100 correlated SNPs for ovules without pollen tubes, full SNP set (EMMA). Chr: Chromosome

rank	p	REMLs	stats	dfs	Chr	Marker	Position
1	8.78045238297807e-10	-224,3108355	6,908653273	84	2	PERL0360737	10635106
2	2.4478551760976e-07	-230,5656335	5,61827746	84	5	PERL0924333	8320952
3	2.6686565630968e-07	-231,0095326	5,597716656	84	1	PERL0107744	12653112
4	7.6242230724182e-07	-232,0278765	5,345242632	84	5	PERL0934238	9128321
5	1.14018329301939e-07	-230,1827468	5,798907207	84	3	PERL0433337	2324104
6	1.14018329301939e-07	-230,1827468	5,798907207	84	5	PERL0880360	3364406
7	1.14018329301939e-07	-230,1827468	5,798907207	84	5	PERL1037546	17701503
8	1.14018329301939e-07	-230,1827468	5,798907207	84	5	PERL1037551	17701740
9	1.39177946749607e-07	-230,376745	5,751978705	84	1	PERL0106534	12539687
10	1.7834237662584e-06	-232,3195683	5,137033818	84	2	PERL0327018	6031993
11	3.0739617218319e-06	-233,3758608	5,00160091	84	1	PERL0100149	12103941
12	3.0739617218319e-06	-233,3758608	5,00160091	84	1	PERL0175236	20798476
13	3.0739617218319e-06	-233,3758608	5,00160091	84	3	PERL0470402	7010320
14	3.21693942925406e-07	-231,1909848	5,553126621	84	4	PERL0846739	17603627
15	5.33472934627802e-07	-230,9254892	5,431676876	84	5	PERL0924322	8320026
16	6.88101658502417e-07	-231,8084337	5,370129244	84	5	PERL0862638	884490
17	6.88101658502417e-07	-231,8084337	5,370129244	84	5	PERL0862649	885664
18	7.3140441585179e-06	-234,2108768	4,782262001	84	1	PERL0044203	5987885
19	7.3140441585179e-06	-234,2108768	4,782262001	84	1	PERL0044214	5990841
20	7.3140441585179e-06	-234,2108768	4,782262001	84	1	PERL0044226	5993156
21	1.05274220542486e-06	-232,3403214	5,266620811	84	1	PERL0268896	30424323
22	1.14016611128123e-06	-232,4175307	5,247101984	84	4	PERL0727239	6513778
23	1.35772053763198e-06	-232,5864783	5,20426058	84	4	PERL0839121	16517649
24	1.35772053763198e-06	-232,5864783	5,20426058	84	4	PERL0839170	16524700
25	1.37843581473123e-06	-232,6011259	5,20053842	84	1	PERL0106529	12539072
26	1.40894724417536e-06	-232,6002064	5,195154587	84	1	PERL0268517	30408327
27	1.50332703309878e-06	-232,6849949	5,179195406	84	4	PERL0808430	12190987
28	1.61527539639852e-06	-232,7544356	5,161490388	84	4	PERL0690403	2823133
29	1.8835470404209e-05	-235,1189289	4,536990066	84	5	PERL0975770	13348145
30	2.22723317260876e-06	-233,064844	5,081956533	84	4	PERL0752005	8013050
31	2.22723317260876e-06	-233,064844	5,081956533	84	4	PERL0845876	17529967
32	2.22723317260876e-06	-233,064844	5,081956533	84	4	PERL0846241	17562143
33	2.37343193077566e-06	-233,1262442	5,066148432	84	4	PERL0672281	1432862
34	2.38354417049474e-06	-231,9198505	5,065090484	84	1	PERL0148825	17541196
35	2.46603255775696e-06	-233,1631954	5,056620761	84	4	PERL0766642	8987279
36	2.46603255775696e-06	-233,1631954	5,056620761	84	4	PERL0766656	8988188
37	2.46603255775696e-06	-233,1631954	5,056620761	84	4	PERL0766713	8993917
38	2.84482906363337e-06	-233,3011135	5,020975005	84	5	PERL0933672	9094854
39	3.53443377077028e-06	-233,5105004	4,966599532	84	5	PERL1122999	26568929
40	3.58084405939148e-06	-233,5230784	4,963322354	84	5	PERL1122964	26566485
41	3.58958630787003e-06	-233,1900426	4,962709678	84	1	PERL0099668	12053242
42	3.80158312646979e-06	-233,5807522	4,948281737	84	5	PERL0879626	3284053
43	3.82005696285871e-06	-232,2959738	4,947061882	84	3	PERL0419782	525776
44	4.38698461661623e-06	-233,7187935	4,912180145	84	4	PERL0809117	12294772
45	4.61721607497984e-06	-233,768074	4,89925601	84	3	PERL0629397	19501879
46	4.73759359554024e-06	-232,9746813	4,892746615	84	5	PERL0924321	8319710
47	4.87469642733669e-06	-233,8203513	4,885526363	84	4	PERL0663604	643962
48	4.96303097402966e-06	-233,8376511	4,880978513	84	2	PERL0294798	2562542
49	5.10317020663032e-06	-233,4643155	4,873922921	84	5	PERL1011507	15963435
50	5.31405588781208e-06	-233,903464	4,863653552	84	5	PERL0943644	9811064
51	5.39485162394063e-06	-233,9179948	4,859823985	84	3	PERL0434258	2478216
52	5.39485162394063e-06	-233,9179948	4,859823985	84	3	PERL0434270	2479813

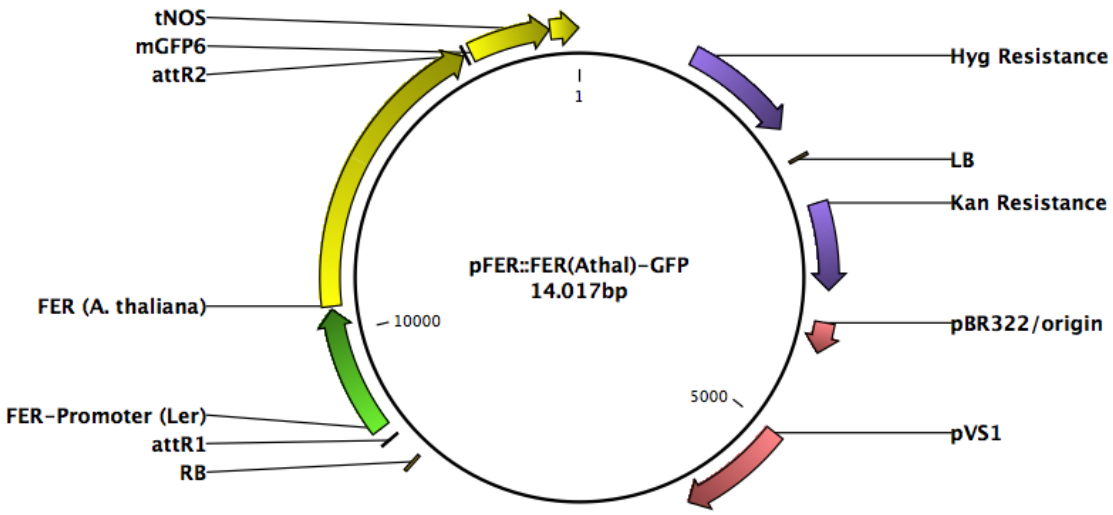
53	6.17507412407329e-06	-234,0480209	4,825477084	84	4	PERL0831942	15460288
54	6.77891706098038e-06	-234,1377893	4,801682968	84	4	PERL0808700	12238277
55	6.91504536517019e-06	-234,1569203	4,796604687	84	1	PERL0260529	29267238
56	6.96939763416707e-06	-233,3520176	4,794604207	84	1	PERL0109073	12866723
57	7.03795309779868e-06	-234,1738653	4,792102513	84	4	PERL0808529	12211760
58	7.03795309779868e-06	-234,1738653	4,792102513	84	4	PERL0809162	12298360
59	7.15237894719343e-06	-234,1893832	4,787979303	84	1	PERL0211511	23791968
60	7.75633431153815e-06	-234,2673443	4,767227203	84	1	PERL0106707	12554313
61	8.24063645183167e-06	-234,3255768	4,751692592	84	1	PERL0052929	7174963
62	8.38509163943318e-06	-233,1992243	4,747230771	84	3	PERL0593246	16175395
63	8.64551104337424e-06	-234,3716777	4,7393728	84	5	PERL0954878	10632459
64	8.72594199359399e-06	-234,3805811	4,736992344	84	4	PERL0725118	6363319
65	8.87385515527859e-06	-233,2178926	4,732669846	84	3	PERL0593337	16177628
66	8.87385515527859e-06	-233,2178926	4,732669846	84	3	PERL0593461	16180091
67	9.68369202972729e-06	-234,2560765	4,710179167	84	3	PERL0484324	8523083
68	9.89047597506854e-06	-233,9706297	4,704729712	84	5	PERL1006527	15558294
69	1.01555331823899e-05	-234,5263546	4,697904299	84	2	PERL0370014	11757030
70	1.04147980195468e-05	-234,5505679	4,691393539	84	4	PERL0666000	841306
71	1.04147980195468e-05	-234,5505679	4,691393539	84	4	PERL0666003	841536
72	1.04928298309889e-05	-234,5577364	4,689464794	84	3	PERL0426011	1363239
73	1.08115423568582e-05	-234,5864732	4,681729132	84	3	PERL0499895	9717225
74	1.13428703102089e-05	-234,6325313	4,669312633	84	5	PERL0933874	9107158
75	1.21279235572148e-05	-234,1520501	4,651964296	84	1	PERL0206921	23467313
76	1.22232138022555e-05	-234,7042887	4,649933244	84	2	PERL0415309	19547995
77	1.22232138022555e-05	-234,7042887	4,649933244	84	2	PERL0415326	19549864
78	1.22232138022555e-05	-234,7042887	4,649933244	84	2	PERL0415333	19550460
79	1.23948823931968e-05	-234,7176765	4,646312739	84	5	PERL0883921	3787697
80	1.28104359352349e-05	-234,749316	4,637746427	84	5	PERL1115724	25660324
81	1.28104359352349e-05	-234,749316	4,637746427	84	5	PERL1115946	25698882
82	1.34911736719423e-05	-234,2116177	4,62428014	84	2	PERL0326990	6026255
83	1.43158219684532e-05	-234,8412807	4,608824049	84	1	PERL0171520	20179578
84	1.46317955679857e-05	-234,8768547	4,603129879	84	1	PERL0142903	17104874
85	1.49386304318338e-05	-234,8967582	4,597713477	84	5	PERL0883685	3745460
86	1.49969264503058e-05	-233,6729275	4,596696623	84	3	PERL0603071	16888879
87	1.62700999434172e-05	-234,7614464	4,575395671	84	5	PERL0883936	3790286
88	1.65405075844614e-05	-234,9944199	4,571080403	84	3	PERL0519254	11110307
89	1.82616913037382e-05	-235,0892871	4,545119213	84	4	PERL0663650	646688
90	1.88387336657443e-05	-235,1190951	4,536944523	84	3	PERL0432098	2147176
91	1.88387336657443e-05	-235,1190951	4,536944523	84	3	PERL0432487	2217317
92	1.88387336657443e-05	-235,1190951	4,536944523	84	3	PERL0433144	2288913
93	1.88387336657443e-05	-235,1190951	4,536944523	84	3	PERL0433170	2292632
94	1.93122314711054e-05	-235,1428719	4,530416018	84	4	PERL0668541	1078676
95	1.94193815338695e-05	-235,1481745	4,528960193	84	4	PERL0839243	16533153
96	1.99611080938295e-05	-235,1745217	4,521717027	84	4	PERL0670551	1260311
97	2.02494367871007e-05	-235,1882605	4,517939238	84	1	PERL0146115	17357928
98	2.07193513612894e-05	-234,8957457	4,511892548	84	1	PERL0258263	28901623
99	2.13201170663435e-05	-234,8598512	4,504352789	84	3	PERL0472548	7255899
100	2.13566251693603e-05	-235,2392334	4,503901263	84	1	PERL0108577	12754087

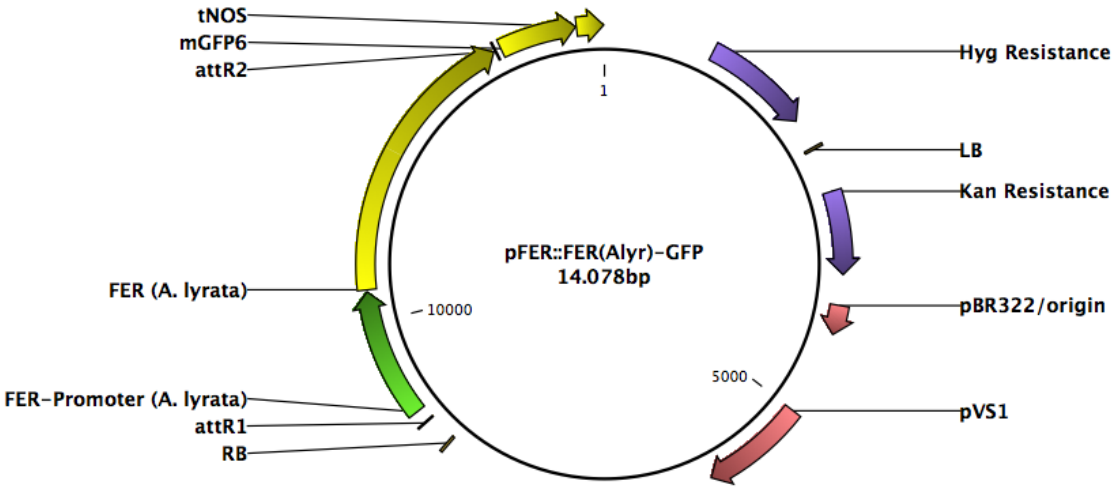
Table A4: Top 100 correlated SNPs for ovules without pollen tubes, SNP0.1 (EMMA). Chr: Chromosome

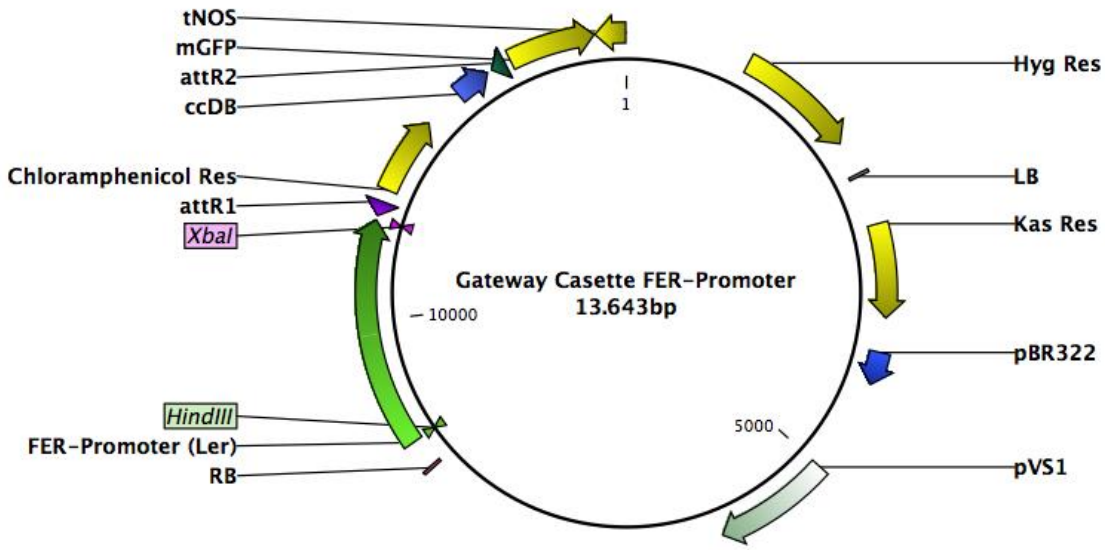
rank	p	REMLs	stats	dfs	Chr	Marker	Position
1	2.66871832805374e-07	-231,0095354	5,597711142	84	1	PERL0107744	12653112
2	7.62452092940856e-07	-232,0278939	5,345233144	84	5	PERL0934238	9128321
3	1.61532795511184e-06	-232,7544467	5,161482361	84	4	PERL0690403	2823133
4	1.72212565286265e-06	-232,3244231	5,145677548	84	2	PERL0327018	6031993
5	2.37351284738279e-06	-233,1262571	5,066139949	84	4	PERL0672281	1432862
6	3.53453501958004e-06	-233,5105083	4,966592336	84	5	PERL1122999	26568929
7	3.5809656161901e-06	-233,5230912	4,963313825	84	5	PERL1122964	26566485
8	3.80173131696073e-06	-233,5807697	4,948271929	84	5	PERL0879626	3284053
9	3.83676255971735e-06	-232,3761484	4,945963731	84	3	PERL0419782	525776
10	4.49341452883568e-06	-233,0027368	4,90612546	84	5	PERL0924321	8319710
11	4.87489868899868e-06	-233,820371	4,885515858	84	4	PERL0663604	643962
12	4.96318903894895e-06	-233,8376616	4,880970446	84	2	PERL0294798	2562542
13	5.31424553794387e-06	-233,9034784	4,863644497	84	5	PERL0943644	9811064
14	6.17529716703115e-06	-234,0480354	4,825467884	84	4	PERL0831942	15460288
15	7.15263520900643e-06	-234,1893977	4,787970141	84	1	PERL0211511	23791968
16	7.3143484034518e-06	-234,2108963	4,782251358	84	1	PERL0044203	5987885
17	7.3143484034518e-06	-234,2108963	4,782251358	84	1	PERL0044214	5990841
18	7.3143484034518e-06	-234,2108963	4,782251358	84	1	PERL0044226	5993156
19	8.03514317633627e-06	-233,2332932	4,758172604	84	3	PERL0593246	16175395
20	8.56581162812731e-06	-233,2625524	4,741752948	84	3	PERL0593337	16177628
21	8.56581162812731e-06	-233,2625524	4,741752948	84	3	PERL0593461	16180091
22	8.64579249905595e-06	-234,3716891	4,739364433	84	5	PERL0954878	10632459
23	8.72619547439094e-06	-234,3805892	4,736984876	84	4	PERL0725118	6363319
24	9.22965426390907e-06	-233,9785047	4,722552675	84	5	PERL1006527	15558294
25	1.04151684855317e-05	-234,550582	4,691384349	84	4	PERL0666000	841306
26	1.04151684855317e-05	-234,550582	4,691384349	84	4	PERL0666003	841536
27	1.13432412657575e-05	-234,6325418	4,669304164	84	5	PERL0933874	9107158
28	1.23951551381022e-05	-234,7176776	4,646307026	84	5	PERL0883921	3787697
29	1.28259237288833e-05	-234,2241647	4,637432399	84	2	PERL0326990	6026255
30	1.49390310357509e-05	-234,896764	4,597706476	84	5	PERL0883685	3745460
31	1.60050586183074e-05	-234,761892	4,579693351	84	5	PERL0883936	3790286
32	1.82624318914597e-05	-235,0893058	4,545108561	84	4	PERL0663650	646688
33	1.93128722965783e-05	-235,1428833	4,530407288	84	4	PERL0668541	1078676
34	1.99618787122623e-05	-235,1745383	4,521706859	84	4	PERL0670551	1260311
35	2.02500879977158e-05	-235,1882714	4,517930764	84	1	PERL0146115	17357928
36	2.03631111009089e-05	-234,8960716	4,516464145	84	1	PERL0258263	28901623
37	2.15581230307723e-05	-235,2482058	4,501422546	84	5	PERL0959465	11038413
38	2.30375202020186e-05	-235,3117396	4,483882675	84	4	PERL0668516	1075168
39	2.38835278264922e-05	-235,346257	4,474336531	84	5	PERL0972436	12918605
40	2.41792431888854e-05	-235,3580315	4,471076835	84	2	PERL0351123	9040817
41	2.57434752767008e-05	-235,4180086	4,454451528	84	4	PERL0750805	7951392
42	2.66255013798254e-05	-235,4502245	4,445503198	84	1	PERL0049699	6806689
43	2.73452120110786e-05	-235,4757312	4,438411658	84	1	PERL0021337	3245775
44	2.73452120110786e-05	-235,4757312	4,438411658	84	1	PERL0021355	3247076
45	2.73452120110786e-05	-235,4757312	4,438411658	84	1	PERL0021361	3248050
46	2.76790633246814e-05	-235,4873344	4,435183229	84	5	PERL0942948	9748275
47	3.04298681537216e-05	-235,577922	4,409931984	84	5	PERL0949406	10170831
48	3.14711599184906e-05	-234,9887814	4,400945941	84	3	PERL0612400	17536683
49	3.24768674865276e-05	-235,6401428	4,392535876	84	5	PERL1034883	17547228
50	3.28202540821477e-05	-235,650191	4,389721942	84	1	PERL0167081	19620570
51	3.43681912318615e-05	-234,7410257	4,377380537	84	3	PERL0509064	10465259
52	3.45084384562558e-05	-235,6981067	4,37628906	84	4	PERL0668505	1074023

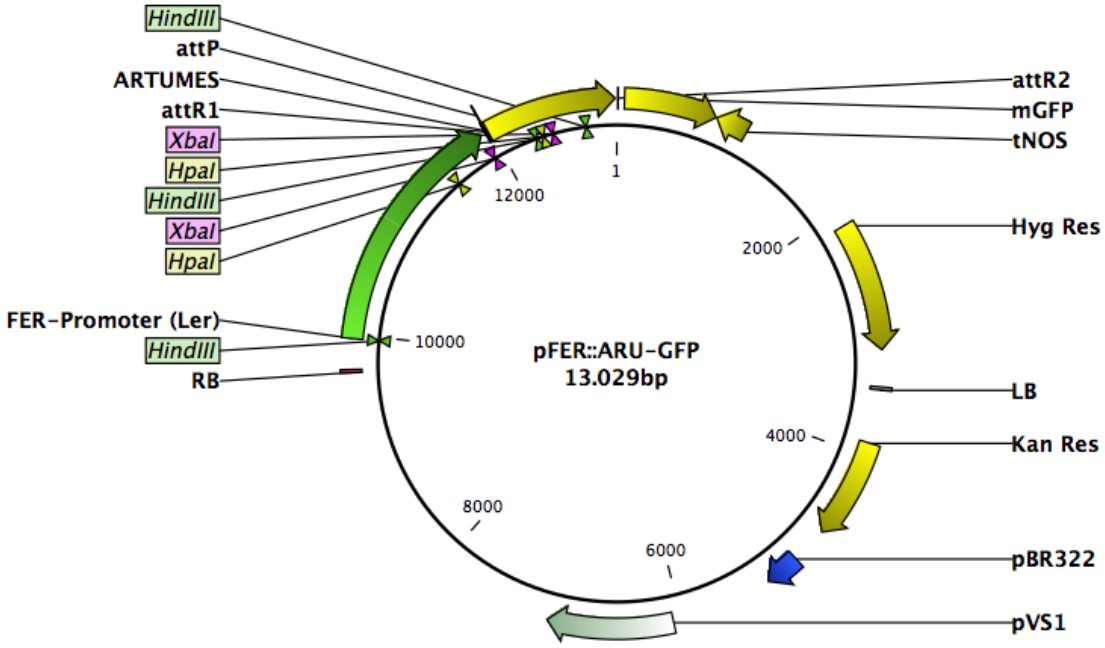
53	3.45084384562558e-05	-235,6981067	4,37628906	84	4	PERL0668523	1075992
54	3.45084384562558e-05	-235,6981067	4,37628906	84	4	PERL0668528	1076842
55	3.53046754443497e-05	-235,71989	4,370172441	84	5	PERL0945923	9943012
56	3.77371117859601e-05	-235,7835239	4,352279957	84	1	PERL0174305	20688369
57	3.7838479118796e-05	-235,2125242	4,351558742	84	5	PERL0905063	6442843
58	3.88159506332982e-05	-235,3520848	4,344698399	84	4	PERL0672280	1432807
59	4.09200805406443e-05	-235,6266994	4,330479921	84	1	PERL0083852	10924678
60	4.10802958749197e-05	-235,8645451	4,329426398	84	5	PERL0968462	12439048
61	4.24844745510889e-05	-235,8966218	4,320359161	84	4	PERL0672065	1409785
62	4.28472887775669e-05	-235,9047386	4,318063404	84	4	PERL0658698	158366
63	4.33988342399028e-05	-235,9169389	4,314609129	84	1	PERL0049708	6808437
64	4.40591157895685e-05	-235,9163199	4,310529161	84	5	PERL1122983	26568088
65	4.42228555470723e-05	-235,9348778	4,309526527	84	3	PERL0656615	23272132
66	4.50268360391845e-05	-235,7980856	4,304654897	84	3	PERL0585962	15708669
67	4.62290560976613e-05	-235,5585118	4,297524611	84	5	PERL0883272	3689218
68	4.62417355068949e-05	-235,9774636	4,297450368	84	1	PERL0021324	3244572
69	4.74541353527899e-05	-235,3782359	4,290440525	84	1	PERL0099405	12035450
70	4.75019616101849e-05	-236,0031107	4,29016756	84	5	PERL1124010	26666023
71	4.79514392573886e-05	-236,0120927	4,287615081	84	2	PERL0348369	8588099
72	5.20242608405401e-05	-235,4264274	4,265485194	84	4	PERL0709782	5086000
73	5.22924879706952e-05	-236,094708	4,264087046	84	4	PERL0829954	15140094
74	5.35386174070991e-05	-236,1171496	4,257680925	84	1	PERL0043985	5954613
75	5.5574925186817e-05	-236,1527271	4,24751581	84	2	PERL0339480	7516005
76	5.61506931414184e-05	-236,0198126	4,244706716	84	3	PERL0597972	16479174
77	6.10183645479803e-05	-236,241726	4,222010271	84	1	PERL0021338	3245838
78	6.12027399817576e-05	-235,3980986	4,221185308	84	5	PERL0863315	979061
79	6.18228518014818e-05	-236,2542016	4,218428167	84	4	PERL0830880	15313545
80	6.19005235587009e-05	-235,9002874	4,218084703	84	5	PERL1122567	26533930
81	6.34551368293776e-05	-236,1712295	4,211296142	84	5	PERL1122620	26536996
82	6.65782809355971e-05	-236,3247567	4,198129432	84	2	PERL0322984	5648594
83	6.73829724113517e-05	-236,3361904	4,194833396	84	3	PERL0610962	17441217
84	6.78175172723857e-05	-236,3423094	4,193069219	84	5	PERL0853164	3392
85	7.03110284307236e-05	-236,3766703	4,183151737	84	1	PERL0044553	6042166
86	7.24480935021092e-05	-235,8998126	4,174917836	84	4	PERL0715994	5602962
87	7.26175971259814e-05	-236,4073798	4,1742748	84	2	PERL0270402	205339
88	7.43497263044566e-05	-236,2303785	4,167785308	84	3	PERL0487283	8905976
89	7.66879519401679e-05	-235,6829559	4,159252126	84	4	PERL0672731	1482517
90	7.69066896886143e-05	-235,4803018	4,15846671	84	4	PERL0700068	4133815
91	7.75660695418444e-05	-236,4700793	4,156112035	84	2	PERL0343862	8047170
92	7.83860288909252e-05	-236,4800744	4,153210647	84	1	PERL0168180	19745717
93	8.21737273790936e-05	-236,2504607	4,140176144	84	2	PERL0345282	8230907
94	8.23334178631276e-05	-236,5267892	4,139639391	84	5	PERL0957580	10837336
95	8.42424129663924e-05	-236,3241138	4,133299263	84	4	PERL0715976	5602110
96	8.52478795002433e-05	-236,5598553	4,130015268	84	2	PERL0339484	7516285
97	8.53061594149693e-05	-236,5605063	4,129826061	84	4	PERL0776608	9602286
98	8.61748887283659e-05	-236,5701249	4,127020363	84	4	PERL0810991	12519934
99	8.65211395419046e-05	-235,7044635	4,125909666	84	3	PERL0521342	11240823
100	8.65755426914642e-05	-235,8748008	4,125735542	84	5	PERL0897895	5748229

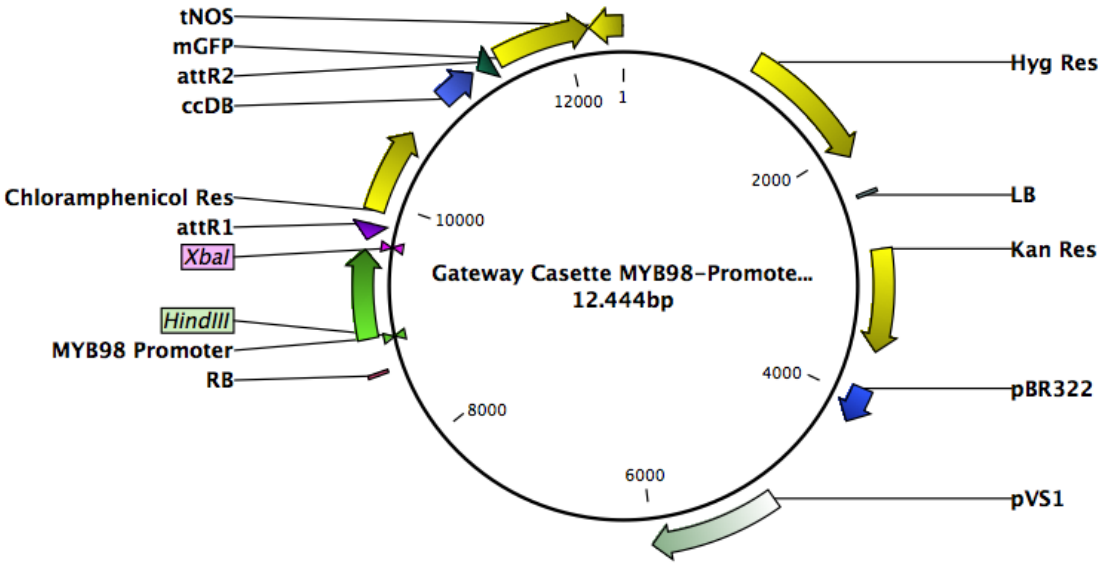
Vector Maps

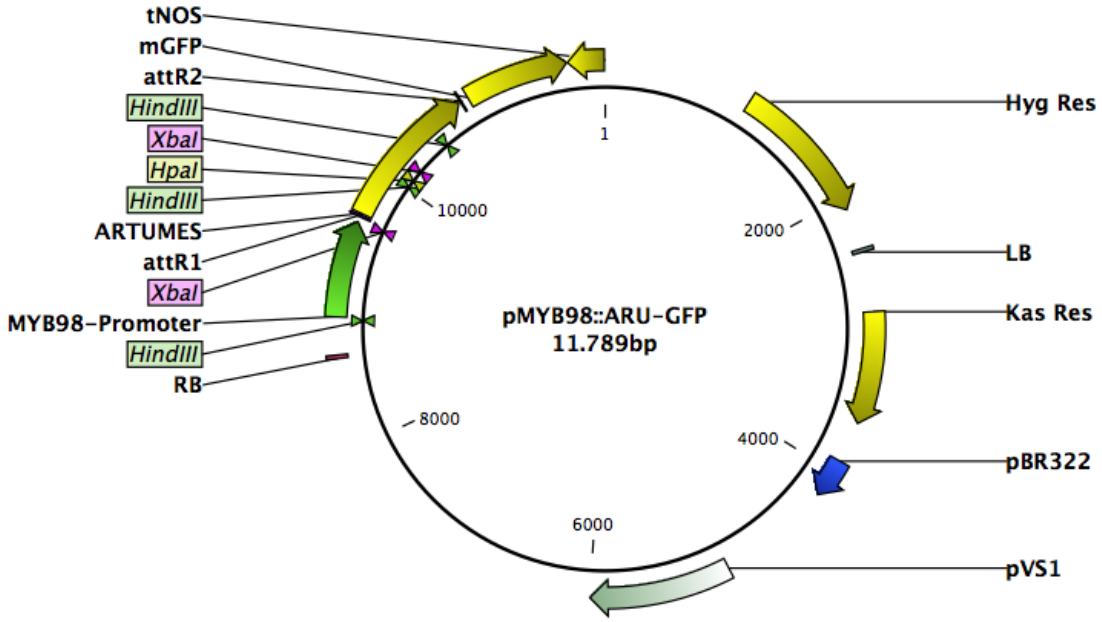

tNOS
mGFP6
attR2
1
Hyg Resistance
LB
Kan Resistance
pFER::FER(Athal)-GFP
14.017bp
FER (A. thaliana)
10000
pBR322/origin
5000
FER-Promoter (Ler)
attR1
RB
pVS1

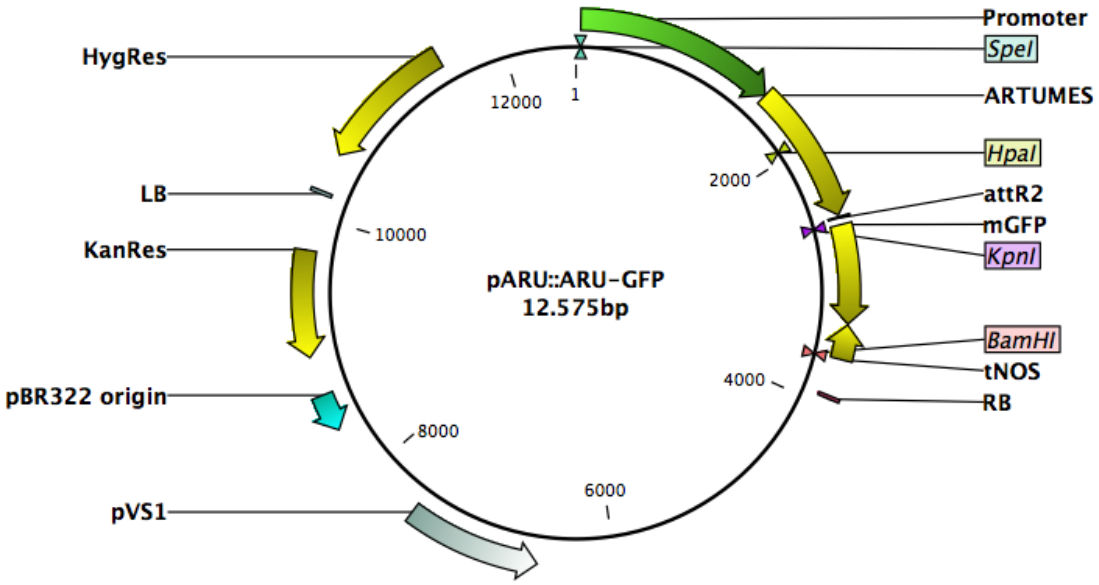

tNOS
mGFP6
attR2
1
Hyg Resistance
LB
Kan Resistance
pFER::FER(Alyr)-GFP
14.078bp
FER (A. lyrata)
10000
pBR322/origin
5000
FER-Promoter (A. lyrata)
attR1
RB
pVS1


tNOS
mGFP
attR2
ccDB
Chloramphenicol Res
attR1
XbaI
HindIII
FER-Promoter (Ler)
RB
Hyg Res
LB
Kas Res
pBR322
pVS1
1
5000
10000
Gateway Casette FER-Promoter
13.643bp


HindIII
attP
ARTUMES
attR1
XbaI
HpaI
HindIII
XbaI
HpaI
FER-Promoter (Ler)
HindIII
RB
attR2
mGFP
tNOS
Hyg Res
LB
Kan Res
pBR322
pVS1
1
2000
4000
6000
8000
10000
12000
pFER::ARU-GFP
13.029bp


tNOS
mGFP
attR2
ccDB
Chloramphenicol Res
attR1
XbaI
HindIII
MYB98 Promoter
RB
Hyg Res
LB
Kan Res
pBR322
pVS1
12000
1
2000
10000
4000
8000
6000
Gateway Casette MYB98-Promote...
12.444bp


tNOS
mGFP
attR2
HindIII
XbaI
HpaI
HindIII
ARTUMES
attR1
XbaI
MYB98-Promoter
HindIII
RB
Hyg Res
LB
Kas Res
pBR322
pVS1
1
2000
10000
4000
8000
6000
pMYB98::ARU-GFP
11.789bp


Promoter
SpeI
ARTUMES
HpaI
attR2
mGFP
KpnI
BamHI
tNOS
RB
HygRes
LB
KanRes
pBR322 origin
pVS1
1
2000
4000
6000
8000
10000
12000
pARU::ARU-GFP
12.575bp